

## PROBING REIONIZATION WITH THE 21 CM GALAXY CROSS-POWER SPECTRUM

ADAM LIDZ<sup>1</sup>, OLIVER ZAHN<sup>2</sup>, STEVEN R. FURLANETTO<sup>3</sup>, MATTHEW MCQUINN<sup>1</sup>, LARS HERNQUIST<sup>1</sup>, AND MATIAS ZALDARRIAGA<sup>1,4</sup>

<sup>1</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>2</sup> Berkeley Center for Cosmological Physics, Department of Physics, University of California, and Lawrence Berkeley National Labs, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>3</sup> Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

<sup>4</sup> Jefferson Laboratory of Physics, Harvard University, Cambridge, MA 02138, USA; [alidz@cfa.harvard.edu](mailto:alidz@cfa.harvard.edu)

Received 2008 June 5; accepted 2008 August 31; published 2008 December 1

### ABSTRACT

The cross-correlation between high-redshift galaxies and 21 cm emission from the high-redshift intergalactic medium promises to be an excellent probe of the Epoch of Reionization. On large scales, the 21 cm and galaxy fields are anticorrelated during most of the reionization epoch. However, on scales smaller than the size of the H II regions around detectable galaxies, the two fields become roughly uncorrelated. Consequently, the 21 cm galaxy cross-power spectrum provides a tracer of bubble growth during reionization, with the signal turning over on progressively larger scales as reionization proceeds. The precise turnover scale depends on the minimum host mass of the detectable galaxies, and the galaxy selection technique. Measuring the turnover scale as a function of galaxy luminosity constrains the characteristic bubble size around galaxies of different luminosities. The cross-spectrum becomes positive on small scales if ionizing photons fail to escape from low-mass galaxies, and these galaxies are detectable longward of the hydrogen ionization edge, because in this case some identifiable galaxies lie outside the ionized regions. The Murchison Widefield Array (MWA) and Low Frequency Array can potentially measure the 21 cm galaxy cross-spectrum in conjunction with mild extensions to the existing Subaru survey for  $z = 6.6$  Ly $\alpha$  emitters. A futuristic galaxy survey covering a sizable fraction of the MWA field of view ( $\sim 800$  deg<sup>2</sup>) can probe the scale dependence of the cross-spectrum, constraining the filling factor of H II regions at different redshifts during reionization, and providing other valuable constraints on reionization models.

*Key words:* cosmology: theory – intergalactic medium – large-scale structure of universe

*Online-only material:* color figures

### 1. INTRODUCTION

Detecting 21 cm emission from the high-redshift intergalactic medium (IGM) will potentially revolutionize our understanding of cosmic reionization. The 21 cm signal promises direct, three-dimensional information regarding the state of the IGM during reionization (e.g., Scott & Rees 1990; Madau et al. 1997; Furlanetto et al. 2006b). Unfortunately, experimental challenges are substantial. In particular, astrophysical foregrounds are expected to be four to five orders of magnitude larger than the signal from the high-redshift IGM. However, known foreground contaminants are spectrally smooth and should be distinguishable from the high-redshift 21 cm signal itself (Zaldarriaga et al. 2004).

Given that we anticipate observational complications, it is important to develop diagnostics to confirm that the detected 21 cm signal indeed originates from the high-redshift IGM. One such approach is to measure the cross-correlation between 21 cm emission and a high-redshift galaxy survey (Furlanetto & Lidz 2007; Wyithe & Loeb 2007). Since most of the anticipated foregrounds come from low-redshift—primarily galactic synchrotron—and not from high-redshift galaxies, the mean 21 cm galaxy cross-power spectrum signal is largely immune to foreground contamination (Furlanetto & Lidz 2007). Detecting a 21 cm galaxy cross-correlation should hence confirm that the detected 21 cm signal comes from the high-redshift IGM. Moreover, continuing efforts are pushing galaxy surveys toward higher redshifts, and it is natural to consider the information that may be gleaned from combining galaxy and 21 cm surveys. Detecting galaxies at very high redshift is extremely challenging

(e.g., Stark et al. 2007; Bouwens et al. 2008), but we will show here that a cross-spectrum detection may already be possible with modest extensions to the Subaru survey (Kashikawa et al. 2006; see also Wyithe & Loeb 2007; Furlanetto & Lidz 2007).

In addition to these important practical advantages, the 21 cm galaxy cross-correlation is potentially sensitive to the size and filling factor of H II regions, the clumpiness of the IGM, and the nature of the ionizing sources. The 21 cm galaxy cross-correlation will also provide a more direct tracer of the interplay between the reionizing sources and the surrounding IGM, than the 21 cm auto-power spectrum. The 21 cm galaxy cross-correlation should hence provide a unique and powerful probe of the Epoch of Reionization (EoR) and early structure formation. Here we follow up on earlier work by Wyithe & Loeb (2007) and Furlanetto & Lidz (2007), and focus on modeling the 21 cm galaxy cross-power spectrum, and exploring the insights that future surveys will provide regarding the EoR.

The outline of this paper is as follows. In Section 2 we establish notation, describe the reionization simulations used in our analysis, and examine the basic simulated signal. We then characterize (Section 3) the dependence of the cross-spectrum on redshift and ionization fraction. In Section 4 we illustrate how the signal is sensitive to the properties of the ionizing sources. In Section 5 we describe its variation with the abundance of Lyman-limit systems. We then examine the signal's dependence on the way in which high-redshift galaxies are selected (Section 6), contrasting the results for Ly $\alpha$ -selected galaxies with galaxies selected through other techniques. In Section 7 we briefly discuss the statistical power of future surveys to constrain reionization through measurements of the

21 cm galaxy cross-power spectrum. Finally, we summarize our main results and conclude in Section 8.

Throughout we consider a  $\Lambda$  cold dark matter (CDM) cosmology parametrized by  $n_s = 1$ ,  $\sigma_8 = 0.8$ ,  $\Omega_m = 0.27$ ,  $\Omega_\Lambda = 0.73$ ,  $\Omega_b = 0.046$ , and  $h = 0.7$ , (all symbols have their usual meanings), consistent with the *Wilkinson Microwave Anisotropy Probe* (WMAP) constraints from Spergel et al. (2007) and Komatsu et al. (2008).

## 2. THE 21 CM GALAXY CROSS-POWER SPECTRUM

In this paper we focus on the cross-power spectrum between the 21 cm and galactic abundance fields. Each field is non-Gaussian and so the cross- and auto-power spectra alone provide an incomplete description of the fields' statistical properties. However, the limited sensitivity of first-generation 21 cm surveys will prohibit detailed imaging of the 21 cm field (McQuinn et al. 2006), and we expect these surveys to have relatively low signal to noise for detecting higher-order moments of the 21 cm field. We hence focus on the cross-power spectrum throughout.

In order to explore the information content of the 21 cm galaxy cross-power spectrum, it is useful to decompose the signal into the sum of several contributing terms. Throughout this work we adopt the limit that the spin temperature of the 21 cm transition is much higher than the cosmic microwave background (CMB) temperature globally (Ciardi & Madau 2003; Pritchard & Furlanetto 2007),  $T_s \gg T_{\text{CMB}}$ , and we ignore peculiar velocities—which should be a good approximation during most of the reionization epoch (Mesinger & Furlanetto 2007). With these assumptions the 21 cm galaxy cross-power spectrum can be written as

$$\Delta_{21,\text{gal}}^2(k) = \tilde{\Delta}_{21,\text{gal}}^2(k)/T_0 = \langle x_{\text{HI}} \rangle [\Delta_{x,\text{gal}}^2(k) + \Delta_{\rho,\text{gal}}^2(k) + \Delta_{x\rho,\text{gal}}^2(k)]. \quad (1)$$

In this equation  $\Delta_{21,\text{gal}}^2(k)$  denotes the cross-power spectrum between the 21 cm field and the galaxy overdensity at wave number  $k = |k|$ . The 21 cm field at spatial position  $r$  is given by  $\delta_T(r) = T_0 \langle x_H \rangle (1 + \delta_x(r))(1 + \delta_\rho(r))$ ;  $T_0$  is the 21 cm brightness temperature, relative to the CMB, at the redshift in question for a fully neutral gas element at the cosmic mean density, and  $\langle x_H \rangle$  is the volume-averaged hydrogenic neutral fraction. The field  $\delta_x(r) = (x_H(r) - \langle x_H \rangle)/\langle x_H \rangle$  is the fractional fluctuation in the neutral hydrogen fraction, while  $\delta_\rho$  is the fractional gas density fluctuation. Similarly  $\delta_g(r) = (n_g(r) - \langle n_g \rangle)/\langle n_g \rangle$  is the fractional fluctuation in galaxy abundance, where  $n_g(r)$  specifies the comoving number density of galaxies at spatial position  $r$ , and  $\langle n_g \rangle$  denotes the volume-averaged galactic abundance. Our notation labels the dimensionless cross-spectrum of two random fields,  $a$  and  $b$ , by  $\Delta_{a,b}^2(k) = k^3 P_{a,b}(k)/(2\pi^2)$  and  $\Delta_{x,\text{gal}}^2$ , for example, is shorthand for the cross-power spectrum between  $\delta_x$  and  $\delta_g$  (and  $P_{a,b}$  is the usual dimensionful cross-spectrum). We use a similar shorthand for  $\Delta_{\rho,\text{gal}}^2(k)$  and  $\Delta_{x\rho,\text{gal}}^2(k)$ . Throughout we work with the power spectrum of the dimensionless field  $\delta_T(r)/T_0$  which we denote by  $\tilde{\Delta}_{21,\text{gal}}^2(k)$ , which is distinguished from the dimensionful power spectrum  $\tilde{\Delta}_{21,\text{gal}}^2(k)$  by the factor of  $T_0$  as in Equation (1).

The individual terms contributing to the 21 cm galaxy cross-power spectrum have the following physical interpretations. The first term,  $\Delta_{x,\text{gal}}^2(k)$ , represents the cross-power spectrum between the neutral hydrogen fraction and galaxy density

fields. The second term,  $\Delta_{\rho,\text{gal}}^2(k)$ , is the cross-power spectrum between the matter and galaxy overdensity fields. The final term,  $\Delta_{x\rho,\text{gal}}^2(k)$ , is a three-field term which would vanish if each contributing field were Gaussian. In fact, we show below that this term is generally significant during reionization (see Lidz et al. 2007a for discussion of a similar term,  $\Delta_{x\rho,\rho}^2(k)$ , which contributes to the 21 cm auto-power spectrum). We will sometimes refer to these three terms respectively as the “ $x$ -gal,” “ $\rho$ -gal,” and “three-field” terms. Let us examine each contributing term from our reionization simulations.

### 2.1. Reionization Simulations

First, we briefly describe the two types of reionization simulations used in this work. The first types are the reionization simulations of McQuinn et al. (2007a). In these simulations, radiative transfer is treated in a post-processing stage using the code of McQuinn et al. (2007b), a refinement of the Sokasian et al. (2001, 2003) code, which in turn uses the adaptive ray-tracing scheme of Abel & Wandelt (2002). The radiative transfer calculation is performed on top of a 130 Mpc/ $h$ ,  $1024^3$  particle dark matter simulation run with an enhanced version of Gadget-2 (Springel 2005). The minimum resolved halo in this simulation is  $\sim 10^{10} M_\odot$ , but smaller mass halos down to the atomic cooling mass (Barkana & Loeb 2001),  $M_{\text{cool}} \sim 10^8 M_\odot$ , are incorporated with the appropriate abundance and clustering as in McQuinn et al. (2007b). Ionizing sources are placed in simulated halos with simple prescriptions. In our fiducial model, we assume that a source's ionizing luminosity is proportional to its host-halo mass. We assume that gas directly traces the dark matter, which should be a good approximation on the large scales of interest here.

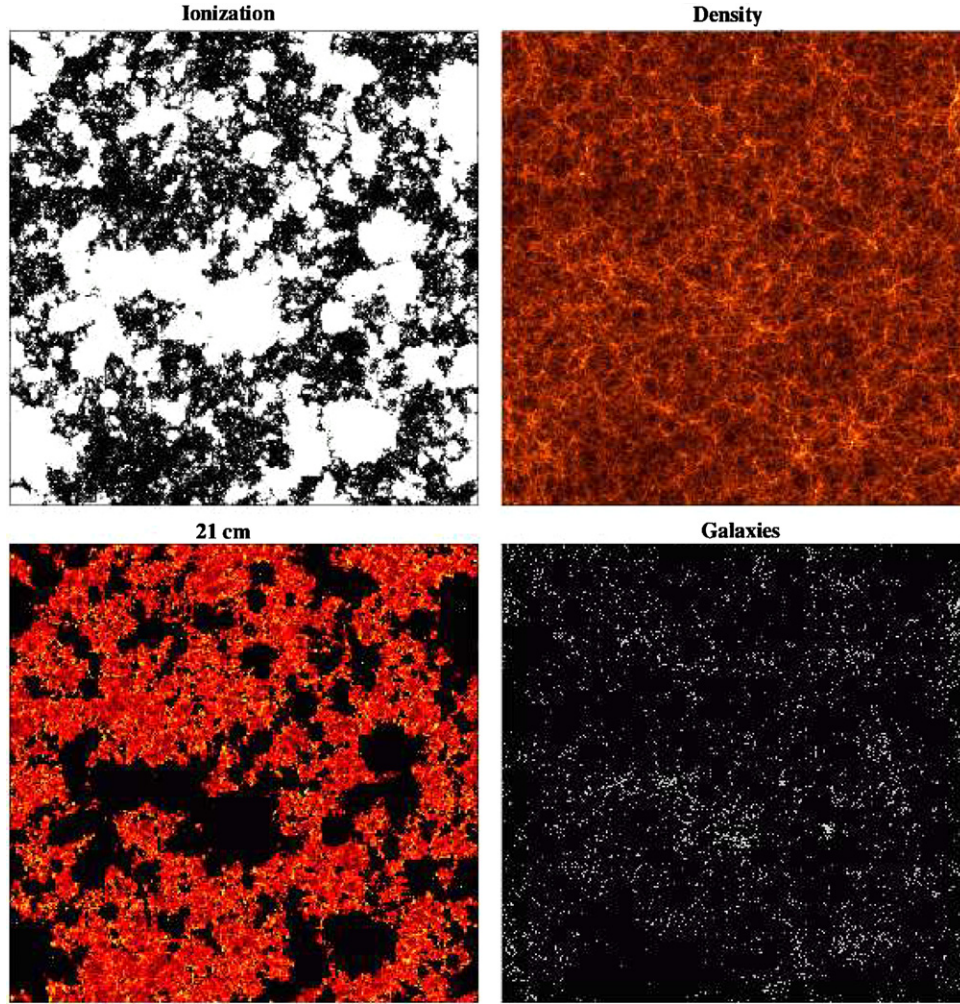
Second, we use an improved version of the hybrid simulation technique of Zahn et al. (2007), which is essentially a Monte Carlo implementation of the analytic model developed by Furlanetto et al. (2004). This technique has the advantage of being extremely fast, while maintaining accuracy. In comparison with the scheme described in Zahn et al. (2007), our present scheme is improved in several ways. First, we use second-order Lagrangian Perturbation Theory (2LPT) to generate realizations of the density field (Crocce et al. 2006) during reionization (as in Lidz et al. 2007b), rather than generating Gaussian random fields. This allows us to incorporate quasilinear effects. Next, we use a scheme similar to that of Mesinger & Furlanetto (2007) to predict the halo distribution from an initial, linear displacement field.

### 2.2. Basic Simulated Signal

Let us now examine the main features of the simulated signal. To begin with, we consider the McQuinn et al. (2007a) simulations, and focus on a model in which all halos down to the atomic cooling mass contain sources with an ionizing luminosity proportional to host-halo mass. Further, we assume that all halos above  $M_{g,\text{min}} = 10^{10} M_\odot$  contain galaxies detectable by our hypothetical survey. In what follows, this prescription for the ionizing sources and the minimum detectable host-halo mass constitutes our fiducial model. We denote the minimum detectable host-halo mass by  $M_{g,\text{min}}$ , and the minimum host-halo mass for the ionizing sources by  $M_{x,\text{min}}$ . Presently we consider a simulation snapshot at  $z = 7.32$ , at which point the filling factor of ionized regions in our model is  $\langle x_i \rangle = 0.54$ .

It is illuminating to inspect the simulated fields visually before calculating their detailed statistics. In Figure 1 we show narrow





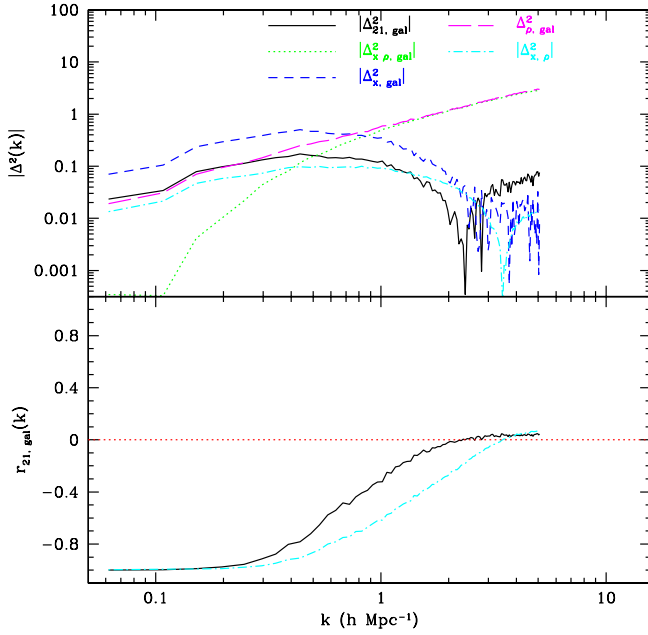
**Figure 1.** Simulated maps of the density, halo, ionization, and 21 cm fields. Each map is 130 Mpc/h on a side and is drawn from a simulation snapshot at  $z = 7.32$  at which point  $\langle x_i \rangle = 0.54$  in our model. The density, ionization, and 21 cm maps are each 1 cell thick (0.25 Mpc/h), while the halo field is from a 60 cell (15 Mpc/h) wedge. On large scales, the bright regions in the overdensity map tend to have more halos, be ionized, and be dim in 21 cm. The correspondence between the bright regions in the halo field, and the dim regions in the 21 cm field, is the signal we characterize and quantify in this paper.

(A color version of this figure is available in the online journal.)

slices through our simulated density, halo, ionization, and 21 cm fields. Here one can clearly see that the bright regions in the halo map correspond to dim regions in the 21 cm map, while dim regions in the halo map correspond to bright regions in the 21 cm map. This anticorrelation is the signal we characterize and calculate in the present paper. As one can see from the panels of Figure 1, the anticorrelation arises because galaxies are more abundant in large-scale overdense regions, which hence ionize before typical regions. As a result, the overdense regions contain less neutral hydrogen during reionization, and emit more dimly in 21 cm than typical regions, while containing more galaxies (see also Wyithe & Loeb 2007).

In order to quantify these visual impressions, we calculate and show the 21 cm galaxy cross-power spectrum in Figure 2. The top panel shows the absolute value of the 21 cm galaxy cross-power spectrum, as well as the individual terms of Equation (1). The bottom panel shows the cross-correlation coefficient between the two fields,  $r(k) = P_{21,\text{gal}}(k)/[P_{21}(k)P_{\text{gal}}(k)]^{1/2}$ . In estimating the cross-correlation coefficient here and throughout this paper, we subtract shot noise from the galaxy power spectrum (before calculating  $r(k)$ ) assuming that it is Poisson—i.e., we assume  $P_{\text{shot}} = 1/n_{\text{gal}}$ , where  $n_{\text{gal}}$  is the abundance of halos above  $M_{\text{g,min}}$ .

The figure reveals several interesting features of the signal. On large scales the 21 cm field is anticorrelated with the galaxy field. As explained and visualized in Figure 1, this occurs because galaxies form first, and ionize their surroundings, in overdense regions. On small scales, the 21 cm and galaxy fields are roughly uncorrelated. We can understand this by examining the small-scale behavior of the constituent terms, as shown in the top panel. The cross-power spectrum between neutral hydrogen fraction and galactic density ( $\Delta_{x,\text{gal}}^2(k)$ , the x-gal term) turns over on small scales, as indicated by the blue-dashed line. This behavior is naturally similar to that of the density-ionization cross-power spectrum, which turns over on scales smaller than the size of the H II regions during reionization (Furlanetto et al. 2004; Zahn et al. 2007). The correlations die off on sub-bubble scales because the entire interior of each H II region is highly ionized, irrespective of the interior density and galaxy fields. For comparison, we additionally plot the cross-power spectrum between neutral hydrogen fraction and matter density. This resembles the cross-power spectrum between neutral hydrogen fraction and galactic density, but it turns over on slightly smaller scales. As we discuss in Section 4 and Section 6.1, the turnover is on smaller scales owing to ionized bubbles around low-mass halos, which host



**Figure 2.** The 21 cm galaxy cross-power spectrum and its constituent terms. The signal is shown for our fiducial model at  $\langle x_i \rangle = 0.54$ . Top panel: the absolute value of the 21 cm galaxy cross-power spectrum (black solid line). The blue dashed line shows the x-gal cross-power spectrum, the magenta long-dashed line shows the  $\rho$ -gal cross-power spectrum, and the green dotted line shows the three-field term. On small scales the three-field and  $\rho$ -gal cross-power spectra cancel each other out rather closely. For contrast, we also show the cross-power spectrum between the neutral hydrogenic fraction and the density field (cyan dot-dashed line). Bottom panel: the cross-correlation coefficient between the 21 cm and galaxy fields as a function of wavenumber. The cyan dot-dashed line indicates the cross-correlation coefficient between the neutral hydrogenic and density fields. The red dotted line indicates the zero-correlation coefficient. The sign of the signal in the top panel can be inferred from the correlation coefficient shown here.

(A color version of this figure is available in the online journal.)

galaxies below the detection threshold of our hypothetical galaxy survey.

The cross-power spectrum between the density field and the galaxy field is shown by the long-dashed magenta line. Note the very strong clustering of these rare galaxies: the cross-power spectrum has an amplitude of unity,  $\Delta_{\rho,\text{gal}}^2(k) \sim 1$ , on a scale of  $k \sim 1.8 h \text{ Mpc}^{-1}$ . On the same scale, the amplitude of the matter power spectrum,  $\Delta_{\rho,\rho}^2(k)$ , is a factor of  $\gtrsim 7$  smaller. Hence even though dark matter clustering is quasilinear on relevant scales, the clustering of detectable host halos may be quite nonlinear on the same scales.

Finally, let us examine the three-field term,  $\Delta_{x\rho,\text{gal}}^2(k)$ . This term is negative in our calculations and appears to closely cancel out the  $\rho$ -gal cross-power spectrum on small scales. Owing to this cancellation, the shape of the 21 cm galaxy cross-power spectrum closely mimics that of the x-gal cross-power spectrum. The 21 cm galaxy cross-correlation may then offer a relatively direct tracer of bubble growth during reionization: it traces the x-gal term, which turns over on scales smaller than that of the H II regions around the minimum mass detectable galaxies. We examine this further in Section 3, but we first pause to consider the three-field term more closely.

In order to understand why the three-field and  $\rho$ -gal terms cancel each other on small scales, it is helpful to combine the two terms into a single one, and consider the two-point correlation function rather than the power spectrum. Here we use similar reasoning to that of Lidz et al. (2007a); see their Section 3.2.

The two terms are combined as

$$\Delta_{\rho,\text{gal}}^2(k) + \Delta_{x\rho,\text{gal}}^2(k) \propto \text{FT}[\langle x(1)\delta_\rho(1)n_g(2) \rangle]. \quad (2)$$

Here 1 and 2 indicate spatial positions  $x_1$  and  $x_2$ , respectively, while FT refers to a Fourier transform. This equation follows from expanding  $x(1)$  on the right-hand side of the equation as  $x(1) = \langle x \rangle(1 + \delta_x(1))$ , and using  $\langle [1 + \delta_x(1)]\delta_\rho(1)n_g(2) \rangle = \langle \delta_\rho(1)n_g(2) \rangle + \langle \delta_x(1)\delta_\rho(1)n_g(2) \rangle$ .

We can write the above two-point function as

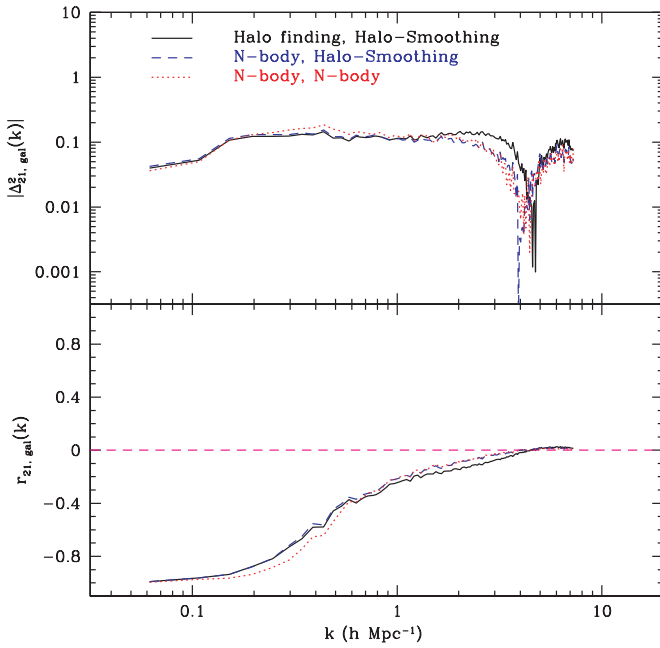
$$\begin{aligned} \langle x(1)\delta_\rho(1)n_g(2) \rangle &= \int dx(1)d\delta_\rho(1)dn_g(2) \\ &\times x(1)\delta_\rho(1)n_g(2)P[x(1), \delta_\rho(1)|n_g(2)]P[n_g(2)]. \end{aligned} \quad (3)$$

Provided we consider separations much smaller than the size of the H II regions, a pair of points (1) and (2) will mostly be either each within the same ionized bubble, or both outside an ionized bubble. If each point is within a bubble then the pixel at position (1) is ionized,  $x(1) = 0$ , and this gives no contribution to the two-point function of Equation (3). On the other hand, spatial points outside the bubbles do not contain detectable galaxies in this model (although see Section 4 for alternate cases),  $n_g(2) = 0$ , and again yield vanishing contributions to the two-point function of Equation (3). The two-point function of Equation (3) must hence vanish on small scales in our fiducial scenario, which explains the cancellation between the  $\rho$ -gal and three-field terms seen in Figure 2. The ionization field is a “mask” that surrounds each galaxy in this model, eliminating the two-point function of Equation (3) on small scales.

### 2.3. Hybrid Simulations

In addition to the full radiative transfer simulations, we use hybrid simulations with two different box sizes in this work: one has  $L_{\text{box}} = 70 \text{ Mpc}/h$ , while the other has  $L_{\text{box}} = 130 \text{ Mpc}/h$ . The density, halo, ionization, and 21 cm fields in each simulation are tabulated on  $512^3$  grid cells. In order to locate the halos using the scheme of Mesinger & Furlanetto (2007), we employ a grid of  $1200^3$  cells in each simulation. The smaller box calculation has higher resolution—resolving halos down to  $M_{\text{min}} \sim 10^8 M_\odot$ —allowing us to accurately identify halos with mass around the atomic cooling mass. The larger box has coarser mass resolution,  $M_{\text{min}} \sim 10^9 M_\odot$ , but better captures the small- $k$  21 cm galaxy cross-spectrum. We refer the reader to Zahn et al. (2007), Mesinger & Furlanetto (2007) for a detailed description and tests of the hybrid scheme. Here we briefly show that estimates of the 21 cm galaxy cross-power spectrum from our hybrid simulations agree well with those from the full radiative transfer simulations.

In order to do this comparison, we use the initial conditions from the McQuinn et al. (2007a)  $N$ -body simulation, and generate the halo field and ionization field using our hybrid scheme. For the purposes of this comparison, in each of our hybrid and radiative transfer calculations, we include ionizing sources only in halos that are well resolved by the  $N$ -body simulation, with  $M_{x,\text{min}} \geq 8 \times 10^9 M_\odot$ . That is, here we do not add low-mass halos into the radiative transfer simulation with the appropriate statistical properties as in McQuinn et al. (2007b) and other sections of this paper. We limit our comparison to masses above  $8 \times 10^9 M_\odot$  because these are the halos directly resolved in our  $N$ -body simulation, before small mass halos are included statistically. We cross correlate the resulting 21 cm field with all halos above our fiducial choice of



**Figure 3.** The comparison between the hybrid and simulated 21 cm galaxy cross-spectrum. Top panel: the absolute value of the 21 cm galaxy cross-power spectrum. Bottom panel: the cross-correlation coefficient between the two fields. In each panel, the red dotted line shows the results from the full radiative transfer simulation. The black solid line shows results from cross-correlating the hybrid 21 cm field with the hybrid halo field. The blue dashed line shows the cross-correlation between the hybrid 21 cm field, and the halo field from the  $N$ -body simulation.

(A color version of this figure is available in the online journal.)

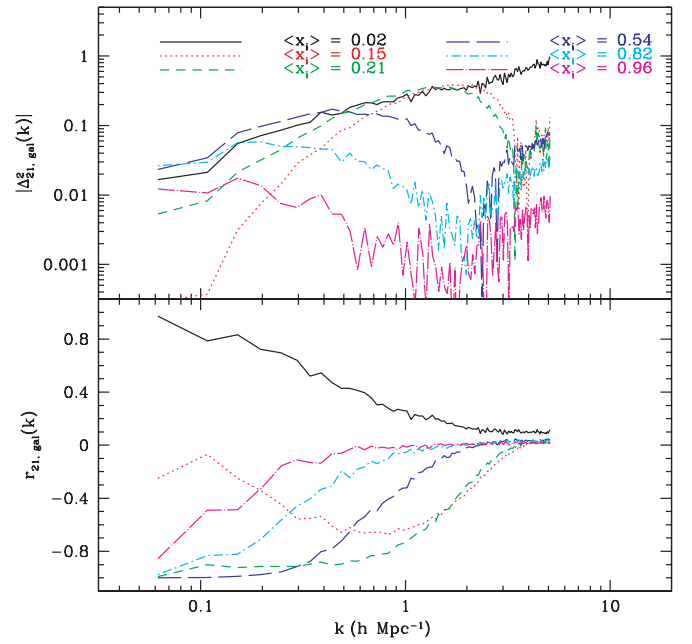
$M_{g, \min} = 10^{10} M_{\odot}$ . The results of this comparison are shown in Figure 3, for outputs with  $\langle x_i \rangle$  just slightly below  $\langle x_i \rangle = 0.5$ .

The agreement between the hybrid and full radiative transfer calculations is quite good. In order to check how much of the small difference between the two calculations comes from differences in the 21 cm field, and how much from differences in the halo fields, we cross correlate the hybrid 21 cm field with the simulated halo field (the blue dashed lines). Differences in the simulated and hybrid halo fields seem to be important on small scales, while differences between the 21 cm fields in the two calculations lead to most of the difference on large scales. Regardless, the hybrid calculations agree well with the full radiative transfer ones, and provide a useful means to estimate the 21 cm galaxy cross-spectrum rapidly.

### 3. REDSHIFT EVOLUTION OF 21 CM GALAXY CROSS-POWER SPECTRUM

Now that we have introduced our simulation tools and understand the basic 21 cm galaxy cross-power spectrum signal; let us examine its dependence on redshift and ionization fraction. How does the signal evolve as the filling factor of H II regions, and their characteristic size, increase? To address this, we calculate the 21 cm galaxy cross-power spectrum from our radiative transfer simulations, considering a wide range of redshifts in order to span most of the reionization epoch. We start by adopting our fiducial model with  $M_{g, \min} = 10^{10} M_{\odot}$  at each redshift for simplicity.

The results of this calculation are shown in Figure 4. At early times, near the very beginning of the reionization process ( $\langle x_i \rangle = 0.02$ ), the galaxy and 21 cm fields are positively correlated on large scales. At this stage, galaxies are extremely rare objects and are only just starting to ionize their



**Figure 4.** The redshift evolution of the 21 cm galaxy cross-power spectrum in our fiducial model. Top panel: the absolute value of the 21 cm galaxy cross-power spectrum at different ionization fractions/redshifts. The redshifts and ionization fractions shown are  $(\langle x_i \rangle, z) = (0.02, 11.46); (0.15, 8.76); (0.21, 8.34); (0.54, 7.32); (0.82, 6.90);$  and  $(0.96, 6.77)$ . Bottom panel: the cross-correlation coefficient between the 21 cm and galaxy fields as a function of wave number.

(A color version of this figure is available in the online journal.)

surroundings. The galaxies turn on first in large-scale overdense regions, which contain more matter and initially more neutral hydrogen than large-scale underdense regions. These regions hence glow more brightly in 21 cm emission and lead to a positive 21 cm galaxy cross-correlation on large scales, as shown by the black solid line in Figure 4.

The galaxies quickly ionize their overdense surroundings, which consequently dim in 21 cm emission. By contrast, the large-scale underdense regions are still mostly free of galaxies and roughly maintain their initial 21 cm brightness temperature. This leads to a brief period where the 21 cm galaxy cross-correlation has low amplitude on large scales, as large-scale overdense regions dim in 21 cm emission and roughly equilibrate in brightness temperature with large-scale underdense regions (Furlanetto et al. 2004; Wyithe & Morales 2007; Lidz et al. 2007b discuss a similar low-amplitude epoch for the 21 cm power spectrum). In our fiducial model, this “equilibration phase” occurs when  $\langle x_i \rangle \sim 0.15$ , as shown by the red dotted line in the figure. This equilibration epoch is relatively brief; the two fields quickly become anticorrelated on large scales.

A caveat to this discussion is that our calculations assume that the spin temperature of the 21 cm transition is globally much larger than the CMB temperature. This approximation will be inaccurate early in the reionization process (Pritchard & Furlanetto 2007; Pritchard & Loeb 2008), and spin temperature fluctuations may complicate the cross-correlation signal close to the equilibration phase. This is beyond the scope of the present paper, but may modify our results at very early times, perhaps when  $\langle x_i \rangle \lesssim 0.1$  (Pritchard & Furlanetto 2007).

More robust, and detectable in the near future, are our results during the bulk of the reionization process, at which point the 21 cm and galaxy fields are anticorrelated. Once



the anticorrelation is established, its scale dependence varies with redshift and ionization fraction. This behavior is shown in the green short-dashed, blue long-dashed, cyan dot-dashed, and magenta dash-dotted lines which span model ionization fractions of  $\langle x_i \rangle = 0.21\text{--}0.96$ . This is the trend discussed in Section 2: galaxies turn on first in overdense environments, and ionize their surroundings. As the ionized fraction increases, and the H II regions grow, the cross-correlation turns over on progressively larger scales, tracing bubble growth during reionization.

We pause here to mention one slight caveat regarding our modeling of the small-scale cross-spectrum. The galactic sources will themselves contain neutral hydrogen, a feature which is not properly included in our calculations. (This leads to a 21 cm signal after reionization; see Wyithe & Loeb 2008; Chang et al. 2008.) This neglected contribution should cause the cross-spectrum to become positive on small scales. Since the signal from the diffuse IGM is much stronger on relevant scales than this galactic contribution (see Lidz et al. 2007b, their Section 2.3, for an estimate), we do not expect this to confuse the determination of the bubble-size-induced turnover.

As remarked previously, the precise turnover scale depends on the minimum host-halo mass of the galaxies observed by our hypothetical survey. Roughly speaking, the turnover is set by the size of ionized regions around detectable galaxy hosts, and is insensitive to the size of ionized regions around fainter galaxies (see Section 6.1). In practice, the minimum detectable host mass—which impacts the turnover scale—may vary with redshift in a complicated way, depending on the flux limit of the survey and the correlation between luminosity and halo mass. This will make the evolution of the 21 cm galaxy cross-spectrum more complicated than the illustrative results of Figure 4, which are at fixed minimum host mass. In a flux-limited survey, the turnover scale will generally evolve less strongly with time: in this case, one detects only massive galaxies at early times, which tend to reside in larger bubbles than average. In order to disentangle the impact of varying minimum host mass and that of varying bubble size and ionization fraction, one could cross correlate the 21 cm signal with galaxies of varying luminosity and use the galaxy autospectrum and luminosity function to help understand the correlation between galaxy luminosity and host-halo mass.

The 21 cm autospectrum itself evolves as the filling factor of H II regions increases. Lidz et al. (2007b) explored how one might use the redshift evolution of the autospectrum to constrain the evolution of the H II region filling factor. The redshift evolution of the cross-spectrum, as considered here, would provide a complementary and essentially independent means for constraining H II region growth. Ultimately, combining the two measurements should provide a cross check on each measurement and increase constraining power. More important, the cross-spectrum provides a much more direct indicator of the characteristic bubble size than the autospectrum. Furthermore, as we show in Section 6.1, measuring the turnover scale as a function of galaxy luminosity allows one to determine the characteristic size of ionized bubbles as a function of luminosity. Note that the ionized regions form under the collective influence of many individual galaxies, but one still expects a statistical trend of the bubble size with galaxy luminosity: more luminous galaxies tend to live in more massive halos, which inhabit larger overdensities, and are typically surrounded by larger ionized regions. Measuring the turnover in the cross-spectrum for different galaxy luminosity bins offers a unique means

of quantifying this trend: this is information which is not obtainable from the 21 cm autospectrum alone. We will discuss the statistical power of several future surveys to detect the cross-spectrum evolution in Section 7.

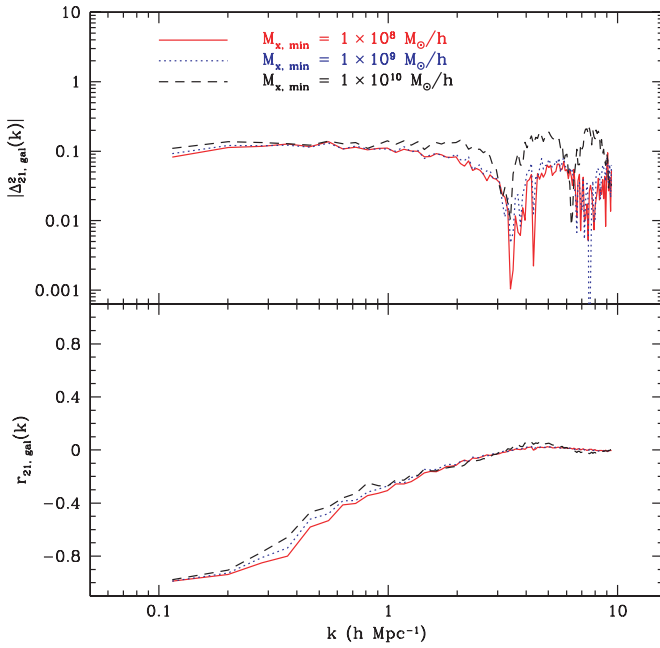
#### 4. DEPENDENCE ON IONIZING SOURCE PROPERTIES

Let us now consider how the 21 cm galaxy cross-spectrum changes with variations around our fiducial model parameters. First, let us examine how the signal depends on the properties of the ionizing sources. Precisely which sources of light produce most of the photons that reionize the IGM is highly uncertain. This depends on many poorly constrained quantities such as the efficiency of star formation as a function of galaxy mass, the high-redshift stellar initial mass function (IMF), the fraction of ionizing photons that escape host galaxies to ionize the IGM and its dependence on host mass, the degree to which photoionization and supernova feedback suppress star formation in low-mass halos, and other factors (e.g., Furlanetto et al. 2006b). A promising route to constrain some of these uncertain parameters is to study the differing impact these sources have on the surrounding IGM. Put simply, the IGM may provide a valuable laboratory for studying the first luminous sources. In this section we show that measurements of the 21 cm galaxy cross-spectrum may help constrain ionizing source properties.

To explore this, let us start with a simple model and vary two of our model parameters. First, we vary  $M_{x,\min}$ , the minimum mass of halos that host sources contributing to reionization. Next, we vary  $M_{g,\min}$ , the minimum host mass of galaxies detectable by our hypothetical galaxy survey. We explore the impact of varying these parameters using 70 Mpc/h hybrid simulations, each normalized to  $\langle x_i \rangle = 0.5$  at  $z = 6.9$ . To begin with, we fix the parameter  $M_{g,\min}$  at  $10^{10} M_\odot/h$  and consider  $M_{x,\min} = 10^8 M_\odot/h$ ,  $10^9 M_\odot/h$ , and  $10^{10} M_\odot/h$  respectively.<sup>5</sup> These are clearly simplified models, but they suffice to illustrate the basic sensitivity of the signal to ionizing source properties. These models should approximate scenarios in which photoheating (Thoul & Weinberg 1996; Navarro & Steinmetz 1997; Dijkstra et al. 2004) or supernova feedback (e.g., Springel & Hernquist 2003) limit the efficiency of star formation in small mass halos and diminish the contribution of these halos to reionization.

Varying  $M_{x,\min}$  across the range shown in Figure 5 only weakly influences the 21 cm galaxy cross-spectrum. On large scales, the amplitude increases slightly with increasing  $M_{x,\min}$ , since the bias of the ionized regions is larger for larger values of  $M_{x,\min}$ . Note, however, that the small-scale turnover occurs at very similar scales for each  $M_{x,\min}$ . This occurs because, in each case shown here, the minimum detectable galaxy mass is larger than (or equal to)  $M_{x,\min}$ . The cross-correlation is mostly insensitive to the bubble sizes around smaller mass, undetectable sources. As alluded to earlier, the turnover in the cross-spectrum depends on the bubble sizes around galaxies above the minimum mass detectable by our hypothetical galaxy survey, and is mostly insensitive to the bubble sizes around lower-mass hosts. Note that the autospectra of the ionization and 21 cm fields do depend on  $M_{x,\min}$  (Furlanetto et al. 2006a, McQuinn et al. 2007b, Lidz et al. 2007b)—models with larger  $M_{x,\min}$  have larger bubbles (on average) at a given  $\langle x_i \rangle$ . However, it appears that the bubble

<sup>5</sup> Note that we generally quote masses in units of  $M_\odot$ , but here and in Section 6.1 (owing to imperfect planning) we use  $M_\odot/h$  units, and so the choice of  $M_{g,\min} = 10^{10} M_\odot/h$  is slightly different from our fiducial choice of  $M_{g,\min} = 10^{10} M_\odot$ .



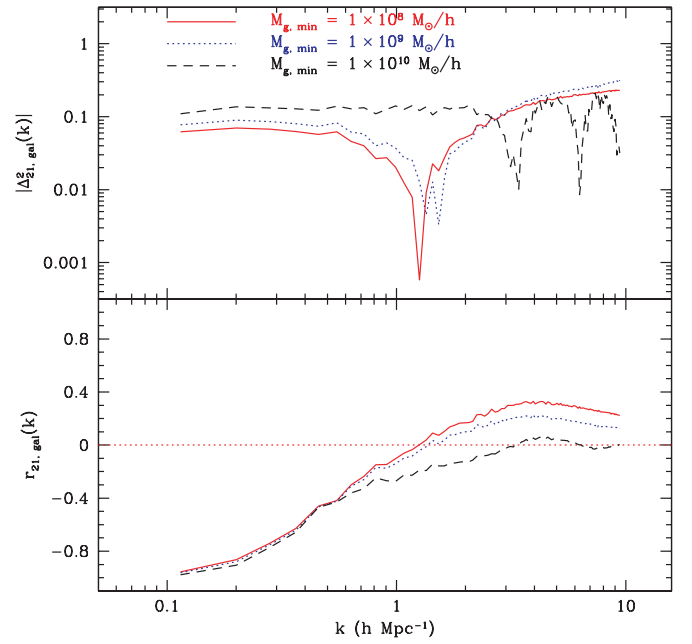
**Figure 5.** The 21 cm galaxy cross-power spectrum for models of varying  $M_{x,\min}$ . In each model curve the efficiency of the ionizing sources is adjusted to yield  $\langle x_i \rangle \sim 0.5$  at  $z = 6.9$ , and the minimum detectable galaxy host has a mass of  $M_{g,\min} = 10^{10} M_{\odot}/h$ . Top panel: the absolute value of the 21 cm galaxy cross-power spectrum. Bottom panel: the correlation coefficient between the galaxy and 21 cm fields in each case. The dependence of the cross-power spectrum on the host-halo mass of the ionizing sources is rather mild.

(A color version of this figure is available in the online journal.)

sizes around high-mass galaxies (with  $M \gtrsim M_{g,\min}$ ) change only slightly with increasing  $M_{x,\min}$ , and hence the turnover scale in the cross-spectrum is insensitive to  $M_{x,\min}$ . We have verified this explicitly by calculating the average ionization as a function of distance around halos with  $M = M_{g,\min}$ , for each of  $M_{x,\min} = 10^8 M_{\odot}/h$  and  $M_{x,\min} = 10^{10} M_{\odot}/h$ . The ionization profiles around the massive halos are very similar in these models, supporting our interpretation.

Another possibility is that ionizing photons escape efficiently only from high-mass galaxies, and that low-mass sources do not contribute to reionizing the IGM. This possibility is, in fact, suggested by the recent escape fraction simulations of Gnedin et al. (2008). Even if low-mass galaxies have a negligible escape fraction, they may still form stars efficiently and be detectable at wavelengths longward of the hydrogen ionization edge. This scenario produces an interesting signature in the 21 cm galaxy cross-power spectrum, provided one has a galaxy survey capable of detecting the, presumably faint, sources in these low-mass halos. In order to explore this, we fix the minimum host-halo mass of sources contributing to the reionization of the IGM at  $M_{x,\min} = 10^{10} M_{\odot}/h$  and calculate the 21 cm galaxy cross-spectrum with a galaxy survey probing sources in host-halo masses larger than each of  $M_{g,\min} = 10^8, 10^9$ , and  $10^{10} M_{\odot}/h$ .

The results of this calculation are shown in Figure 6. On large scales, the amplitude of the cross-spectrum increases as one raises the minimum detectable host-halo mass. This increase simply owes to the usual increase in galaxy bias with increasing minimum host-halo mass. Perhaps more interesting, however, are the results on small scales when the minimum detectable host-halo mass is lower than the minimum ionizing source mass. In this case (see the model curves with  $M_{g,\min} = 10^8$

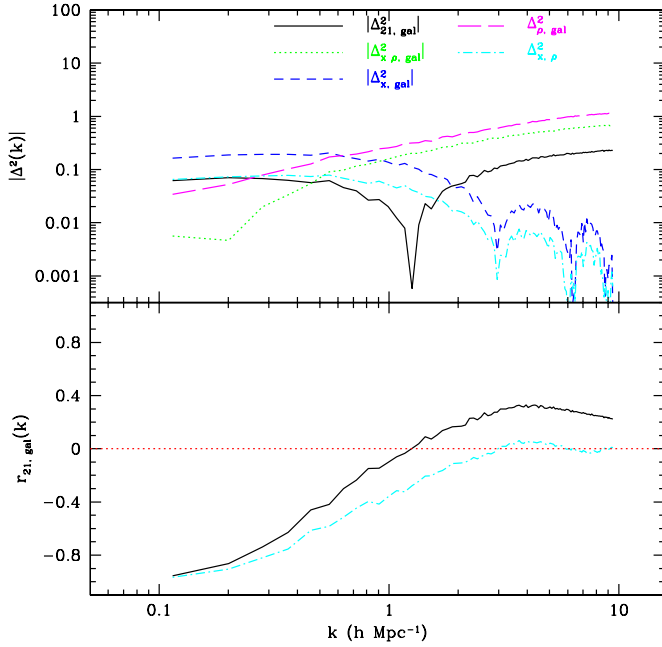


**Figure 6.** 21 cm galaxy cross-power spectra for models where sources in high-mass halos produce most of the ionizing photons. In each model, the minimum host-halo mass of sources that allow ionizing photons to escape into the IGM is  $M_{x,\min} = 10^{10} M_{\odot}/h$ . Top panel: the absolute value of the 21 cm galaxy cross-spectrum for galaxy surveys with minimum detectable host-halo masses of  $M_{g,\min} = 10^8, 10^9$ , and  $10^{10} M_{\odot}/h$ . Bottom panel: the correlation coefficient between the galaxy and 21 cm fields in each case. The cross-spectrum and correlation coefficient turn positive on small scales for cases in which the galaxy survey detects sources with mass below the minimum mass ionizing source.

(A color version of this figure is available in the online journal.)

and  $10^9 M_{\odot}/h$ ), the cross-spectrum turns over on larger scales than in the model in which  $M_{g,\min} = M_{x,\min}$ , it then reverses sign, and goes positive on small scales. Detecting this behavior would indicate that ionizing photons escape only from massive host halos, and not from lower-mass hosts. An ambitious survey is needed to detect the faint sources in low-mass halos, and to detect the 21 cm galaxy cross-spectrum on small scales (see Section 7). Nonetheless, the proposed signature would provide an interesting indication of a small escape fraction from low-mass galaxies. Moreover, this signature is relatively direct—any indication of a small escape fraction from low-mass galaxies in the 21 cm autospectrum will be more subtle, and likely degenerate with other effects. Note, however, that neutral hydrogen in the galaxies themselves may also result in a positive small-scale cross-spectrum (Section 3), and it might be tricky to distinguish this from our escape fraction scenario.

In order to understand this effect better, it is useful to calculate each of the terms in Equation (1) separately. In Figure 7 we examine each of these terms for our model with  $M_{x,\min} = 10^{10} M_{\odot}/h$  and  $M_{g,\min} = 10^8 M_{\odot}/h$ . In this case, the three-field and  $\rho$ -gal terms do not cancel each other, unlike in our fiducial case (Figure 2). This occurs because in this model low-mass galaxies do not leak ionizing photons into the IGM, and can hence reside outside the ionized regions which—in this model—are formed only by sources residing in higher-mass halos. In this way, some low-mass halos escape the “masking” effect of the ionized regions (see Section 2.2 and Equation 3), and—since these low-mass galaxies are correlated with the underlying density field—produce a positive small-scale 21 cm galaxy cross-power spectrum.



**Figure 7.** The decomposition of the 21 cm galaxy cross-spectrum for a model in which the minimum detectable galaxy host mass is below the minimum host mass for ionizing sources. Here we show the decomposition for a model with  $M_{x,\min} = 10^{10} M_\odot/h$  and  $M_{g,\min} = 10^8 M_\odot/h$ . Top panel: as in Figure 2, the absolute value of the 21 cm galaxy cross-spectrum, as well as the x-gal,  $\rho$ -gal, and three-field terms. Additionally, we show the cross-power spectrum between the neutral hydrogen field and the density field itself. Unlike in our previous models, the three-field and  $\rho$ -gal terms do not perfectly cancel each other on small scales. This results because there are detectable galaxies outside ionized regions in this model. Consequently, the 21 cm galaxy cross-spectrum changes sign around  $k \sim 1 h \text{ Mpc}^{-1}$  and goes positive on small scales. Bottom panel: the cross-correlation coefficient between the 21 cm and galaxy fields, as well as the cross-correlation between the neutral hydrogen and density fields.

(A color version of this figure is available in the online journal.)

## 5. THE IMPACT OF LYMAN-LIMIT SYSTEMS

Next we consider the impact of Lyman-limit systems on our 21 cm galaxy cross-spectrum calculations. Once reionization is complete, most ionizing photons are absorbed in dense blobs of neutral gas known as Lyman-limit systems. Lyman-limit systems can also limit the mean free path of ionizing photons and halt the growth of H II regions (Furlanetto & Oh 2005; McQuinn et al. 2007b) during reionization itself, particularly toward the end of reionization. The precise physical nature and abundance of these systems at high redshift is highly uncertain, as is their role as photon sinks during reionization. Lyman-limit systems may be especially numerous and have a strong effect if “mini-halos”—halos with mass less than the atomic cooling mass—manage to survive preheating prior to reionization (Oh & Haiman 2003) and are abundant during reionization (Haiman et al. 2001; Barkana & Loeb 2002; Shapiro et al. 2004).

In order to quantify the impact of Lyman-limit systems on the 21 cm galaxy cross-spectrum, we use the hybrid simulation scheme of Zahn et al. (2007), generalized to include the recombination excursion-set barrier of Furlanetto & Oh (2005). In order to capture the small- $k$  power spectrum—where we expect the Lyman-limit systems to have the most impact—we use the  $L_{\text{box}} = 130 \text{ Mpc}/h$  hybrid simulation. In this section, for simplicity, our hybrid simulation adopts the pure Press–Schechter ionization barrier of Furlanetto et al. (2004), rather than the halo-smoothing algorithm (see Zahn et al. 2007 for comparisons). Since the present work is the first to incorporate

the recombination barrier into a hybrid simulation scheme, we briefly review this model here, but refer the reader to Furlanetto & Oh (2005) for more details. The recombination barrier reflects the requirement that for an H II region to grow, the instantaneous rate of photon production from the sources within the H II region must at least match the recombination rate of the ionized material inside the H II region.

In Furlanetto & Oh (2005), the recombination rate is calculated using the model of Miralda-Escudé et al. (2000). In this model at any given time, the interior of an H II region is ionized up to islands of small-scale overdensity  $\Delta_i$ , above which the gas is neutral. The mean free path to ionizing photons is then determined by the volume-filling factor of these overdense islands. In particular, the proper mean free path to ionizing photons is given by

$$\lambda(z) = \lambda_0(z) [1 - F_v(\Delta_i)]^{-2/3}, \quad (4)$$

where  $F_v(\Delta_i)$  denotes the volume-filling factor of regions with  $\Delta < \Delta_i$ , and  $\lambda_0(z)$  is a normalization factor, which is given by  $\lambda_0(z)H(z) = 60 \text{ km s}^{-1}$  in Miralda-Escudé et al. (2000). Here we leave  $\lambda_0(z)$  as a free parameter to gauge the dependence of our results on the observationally and theoretically uncertain mean free path. The filling factor  $F_v(\Delta_i)$  is computed using the gas density pdf in Miralda-Escudé et al. (2000). Similarly, the recombination rate for the ionized gas, in a region of large-scale overdensity  $\delta$  ionized up to an overdensity  $\Delta_i$ , is given by:

$$A(\delta, \Delta_i) = \alpha_A n_e (1 + \delta) \int_0^{\Delta_i} d\Delta \Delta^2 P(\Delta). \quad (5)$$

Here  $\alpha_A$  denotes the case-A recombination coefficient for the ionized gas, which we assume to be at  $10^4 \text{ K}$ ,  $n_e$  denotes the mean electron density in the IGM, and  $P(\Delta)$  is the gas density pdf from Miralda-Escudé et al. (2000). We assume helium is mostly singly ionized, but not doubly ionized, within the bubble interiors. The recombination rate formula assumes that the density probability distribution function (pdf),  $P(\Delta)$ , is independent of large-scale overdensity,  $\delta$ , which should be a good approximation for the large scales relevant here (Furlanetto & Oh 2005).

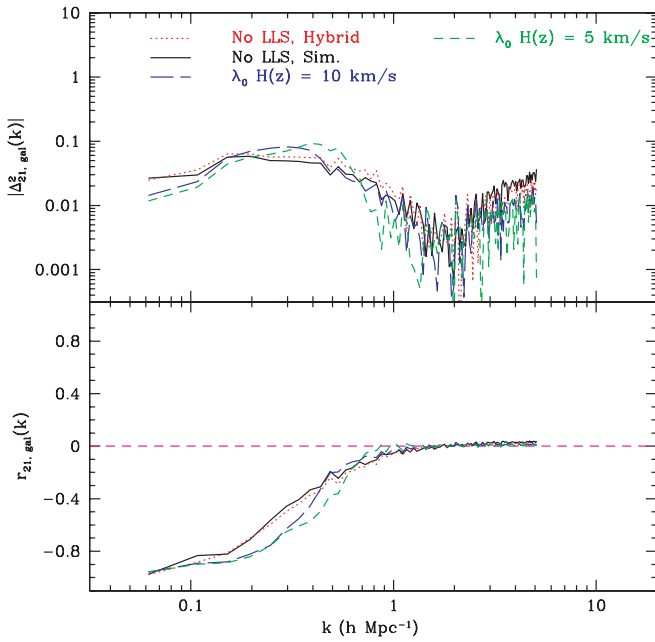
With this formula for the recombination rate in hand, Furlanetto & Oh (2005) write an excursion set barrier for a region of size  $R$  and overdensity  $\delta$  to overcome recombinations and be ionized by interior sources. In our notation, this formula is

$$\zeta \frac{df_{\text{coll}}(\delta, R)}{dt} > A(\delta, R), \quad (6)$$

where  $\zeta$  denotes the ionizing efficiency of the sources,  $f_{\text{coll}}$  denotes the collapse fraction in halos above the minimum host-halo mass, and  $R$  is equated with the mean free path,  $\lambda$ , which sets  $\Delta_i$  through Equation (4), and  $A(\delta, R)$  through Equation (5). We implement this barrier, and apply it in a Monte Carlo fashion (Zahn et al. 2007; Mesinger & Furlanetto 2007), in conjunction with the normal Furlanetto et al. (2004) barrier. This barrier effectively prohibits ionizing photons from propagating long distances, as regulated by the parameter  $\lambda_0(z)$ , and decreases the level of ionization fluctuations on large scales.

In order to see how this can impact the 21 cm galaxy cross-power spectrum, we calculate the signal with hybrid simulations of varying  $\lambda_0(z)$ . In particular, we consider  $\langle x_i \rangle = 0.8$  and vary  $\lambda_0(z)$  over the range  $\lambda_0(z)H(z) = 5 - 60 \text{ km s}^{-1}$ . We span here a rather broad range of models, which is appropriate given





**Figure 8.** The dependence of the 21 cm galaxy cross-power spectrum on the abundance of Lyman-limit systems. Top panel: the absolute value of the cross-spectrum. Bottom panel: the cross-correlation coefficient. The black solid line shows the cross-spectrum in our fiducial model at  $\langle x_i \rangle = 0.8$  ( $z = 6.90$ ). The red dotted line shows an equivalent model from our hybrid simulation scheme. The green short-dashed and blue long-dashed lines show cross-spectrum calculations for models with abundant Lyman-limit systems (see the text). The recombination barrier scales (see the text) for the models with  $\lambda_0(z)H(z) = 5 \text{ km s}^{-1}$ , and  $\lambda_0(z)H(z) = 10 \text{ km s}^{-1}$  are  $R_{\text{rec, barr}} = 4, 7 \text{ Mpc/h}$  respectively. The Lyman-limit systems force the cross-spectrum to turn over toward large scales, but the effect is relatively mild.

(A color version of this figure is available in the online journal.)

our limited observational constraints on the mean free path to ionizing photons at high redshift. The results of calculations with no Lyman-limit systems, and each of  $\lambda_0(z)H(z) = 5$  and  $10 \text{ km s}^{-1}$ , are shown in Figure 8. Since  $\lambda_0(z)H(z) = 5 \text{ km s}^{-1}$  is 1/12th of the fiducial Miralda-Escudé et al. (2000) value, our values represent rather extreme choices for the mean free path. A more meaningful characterization than  $\lambda_0(z)$ , is the characteristic scale where the recombination barrier (Equation 6) crosses the usual Furlanetto et al. (2004) barrier (see Furlanetto & Oh 2005). We denote the scale where the two barriers cross by  $R_{\text{rec, barr}}$ . For models with  $\lambda_0(z)H(z) = 5, 10 \text{ km s}^{-1}$  (and  $\langle x_i \rangle = 0.8$ ,  $z = 6.90$ ), the barriers cross at respective radii of  $R_{\text{rec, barr}} = 4$  and  $7 \text{ Mpc/h}$ . Note that, for our choice of model parameters, the recombination barriers are not so steep, and so some fraction of points do manage to cross the barriers on smoothing scales roughly twice as large as  $R_{\text{rec, barr}}$ . Comparing the red and black model curves, we see that the hybrid scheme accurately captures the 21 cm galaxy cross-power spectrum signal from our full radiative transfer simulations in the no Lyman-limit system case as in Figure 3.

The model curves with  $\lambda_0(z)H(z) = 10 \text{ km s}^{-1}$  ( $R_{\text{rec, barr}} = 7 \text{ Mpc/h}$ ) and  $\lambda_0(z)H(z) = 5 \text{ km s}^{-1}$  ( $R_{\text{rec, barr}} = 4 \text{ Mpc/h}$ ) illustrate that decreasing the mean free path forces the 21 cm galaxy cross-power spectrum to turn over toward large scales, rather than simply flattening out on large scales as in our fiducial model.<sup>6</sup> The cross-spectrum develops a more

well-defined characteristic scale as the mean free path decreases. This trend results because decreasing the mean free path limits the formation of very large H II regions which in turn reduces the amount of large-scale cross-power. Although the figure illustrates a clear trend of decreasing large-scale power with decreasing mean free path, we caution that there is no simple one-to-one correspondence between mean free path and H II region size. As a specific illustration of the distinction between the mean free path and H II region size, consider the post-reionization IGM. In the post-reionization IGM, essentially the entire volume of the IGM is ionized, and the bubble size hence infinite, while the mean free path is still finite.

Note that although the figure shows a clear trend of decreasing large-scale power with decreasing mean free path, the dependence on the abundance of Lyman-limit systems is rather weak. On the one hand, this may preclude strong constraints on the abundance of Lyman-limit systems from future measurements of the 21 cm galaxy cross-power spectrum. On the other hand, it implies that the mean free path to ionizing photons is not a very important factor in our modeling of the cross-spectrum. This should allow us to robustly constrain other parameters from future measurements, in spite of our ignorance of the high-redshift mean free path.

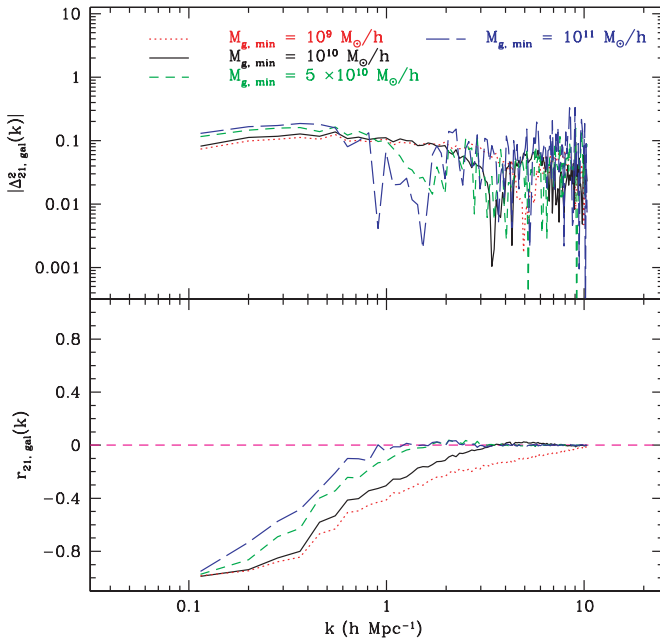
## 6. DEPENDENCE ON THE GALAXY-SELECTION TECHNIQUE

In this section, we consider the dependence of the 21 cm galaxy cross-power spectrum on the manner in which the galaxies are selected. Thus far we have calculated the 21 cm galaxy cross-power spectrum by cross correlating our 21 cm field with all simulated halos above some minimum detectable halo mass cut. In other words, we assume that each simulated dark matter halo contains one luminous galaxy, and that the flux limit of our hypothetical galaxy survey corresponds precisely to a minimum host-halo mass. This is clearly a vast simplification, and so it is important to explore the signal's sensitivity to the minimum mass cut. Note, however, that since we consider only scales much larger than the halo virial radius, we are *not* sensitive to the distribution of galaxies within each host halo (e.g., Scoccimarro et al. 2001). Another important effect is that galaxies selected on the basis of Ly $\alpha$  emission will have a different 21 cm galaxy cross-correlation than galaxies selected by, for example, the Lyman-break technique. We presently explore the sensitivity of our results to the type of galaxy selected by our hypothetical survey.

### 6.1. Minimum Detectable Mass

Let us first fix the population of galaxies responsible for reionizing the IGM, and the resulting ionization field, while varying the minimum host-halo mass containing galaxies detectable by our hypothetical survey. We explored this issue somewhat already in Section 4, but there we focused on scenarios in which ionizing photons do not escape from low-mass galaxies—i.e., cases where  $M_{\text{g, min}} < M_{\text{x, min}}$ . Here we focus on models in which ionizing photons manage to escape from low-mass halos, yet such sources are too faint to be detectable by our hypothetical galaxy survey. In other words, we consider cases where  $M_{\text{g, min}} > M_{\text{x, min}}$ . This is likely the more relevant case for first generation surveys where it will be difficult to detect the presumably faint galaxies that reside in low-mass halos.

<sup>6</sup> In our fiducial model, the 21 cm galaxy cross-power spectrum should turn over on some scale larger than that of our simulation box. Note, however, that foreground contamination may make such scales inaccessible to future 21 cm observations (McQuinn et al. 2006).



**Figure 9.** Dependence of 21 cm galaxy cross-power spectrum on the minimum detectable galaxy mass. Top panel: the cross-power spectrum for a survey that detects all galaxies in halos of mass larger than  $M_{g,\min} = 10^9 M_\odot/h$ ,  $M_{g,\min} = 10^{10} M_\odot/h$ ,  $5 \times 10^{10} M_\odot/h$ , and  $10^{11} M_\odot/h$ , respectively. For each model curve, we fix the minimum host mass of the ionizing sources at  $M_{x,\min} = 10^8 M_\odot/h$ . Bottom panel: the correlation coefficient between the galaxy and 21 cm fields in each case.

(A color version of this figure is available in the online journal.)

Another point is that it is unlikely that all halos above some  $M_{g,\min}$  host galaxies that actively produce detectable photons at any given instant of time. In other words, the “duty cycle”—the fraction of halos above a given mass which contain galaxies actively radiating at a particular time—is likely less than unity. As quantified below, varying the minimum detectable host mass impacts the mean 21 cm galaxy cross-power spectrum, but reducing the duty cycle of detectable galaxies does not by itself change the average cross-power spectrum signal. This is because galaxy bias is independent of duty cycle, provided that the duty cycle is itself independent of mass for halos above the minimum detectable host mass. Decreasing the duty cycle instead increases the level of Poisson fluctuations in the galaxy abundance, which increases the cross-spectrum variance—and makes the cross-spectrum more difficult to detect (Furlanetto & Lidz 2007, Section 7)—while preserving the average cross-power spectrum. Presently, we focus on how the minimum host-halo mass impacts the mean signal, and defer a discussion of the signal variance to Section 7. Of course our assumption that the duty cycle is independent of host mass may be too simplistic and modifying this assumption may impact our results in detail. Our simple model should, however, suffice to illustrate the basic sensitivity to host-halo mass. Moreover, in practice one can constrain the run of duty cycle with halo mass from the observed galaxy luminosity function and galaxy–galaxy auto-power spectrum, which can then inform models for the 21 cm galaxy cross-spectrum.

The results of varying the minimum detectable host-halo mass are shown in Figure 9. Here we use the 70 Mpc/h hybrid simulation, fix  $M_{x,\min} = 10^8 M_\odot/h$  (just a little above the atomic cooling mass), and vary  $M_{g,\min}$  from  $10^9 M_\odot/h$  to  $10^{10} M_\odot/h$ ,  $5 \times 10^{10} M_\odot/h$ , and  $10^{11} M_\odot/h$ . On large scales one

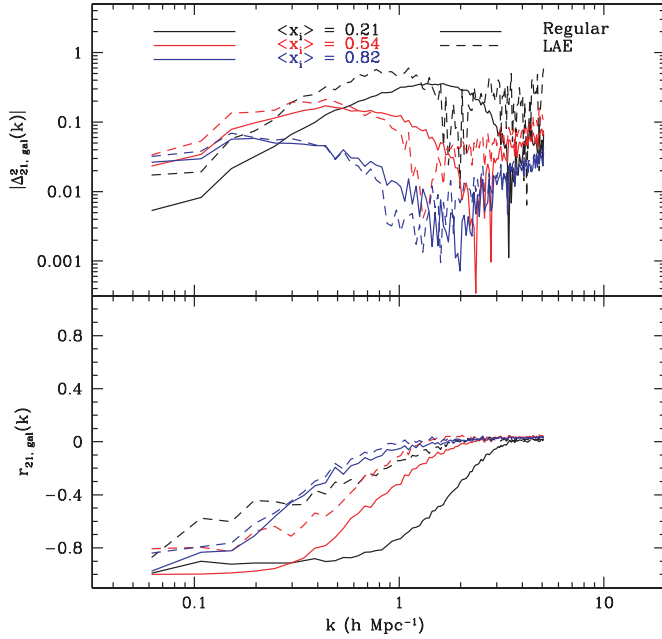
sees the usual increase in the amplitude of the cross-spectrum as  $M_{g,\min}$  increases. On small scales, the cross-spectrum turns over on progressively smaller scales as  $M_{g,\min}$  decreases, and the cross-correlation starts to sample the small bubbles around the lower-mass halos. This is a continuation of the behavior seen in Figure 5. It illustrates that the turnover scale needs to be interpreted with caution, since it is sensitive to  $M_{g,\min}$ . The dependence of turnover scale on luminosity is, however, very interesting; examining it amounts to a measurement of the characteristic bubble size around galaxies of varying luminosity, as anticipated in previous sections. We reiterate that this trend highlights the scientific benefit of measuring the 21 cm galaxy cross-spectrum, as is impossible to determine the luminosity dependence of the bubble size distribution from the 21 cm autospectrum alone.

## 6.2. Ly $\alpha$ -Selected Galaxies

A successful approach for finding high-redshift galaxies is to search for Ly $\alpha$  emission, which is frequently strong in young galaxies (Partridge & Peebles 1967). There are numerous existing and planned Ly $\alpha$  emitter (LAE) surveys (e.g., Rhoads et al. 2004; Kashikawa et al. 2006; Stark et al. 2007), with the Subaru telescope currently providing the largest high-redshift sample, consisting of  $\sim 58$  photometric LAEs at  $z = 6.5$  discovered in a  $\sim 30' \times 30'$  field (Kashikawa et al. 2006). LAE surveys have an advantage over high-redshift Lyman-break surveys in that they target narrow wavelength intervals, in between strong night sky background lines, in search of strong emission lines. This allows one to detect galaxies that are unobservable by Lyman-break selection owing to the strong night sky background at the relevant wavelengths; sizable Lyman-break galaxy catalogs at very high redshift likely await a wide field, near-infrared instrument in space or 30 m class telescopes on the ground. Existing LAE surveys and their extensions hence likely provide the first opportunity to detect the 21 cm galaxy cross-power spectrum, particularly if the IGM is partly neutral at  $z \sim 6.6$  (Wyithe & Loeb 2007; Furlanetto & Lidz 2007, Section 7).

To this end, we would like to consider the cross-correlation between 21 cm and Ly $\alpha$ -selected galaxies. In contrast to galaxies selected via, for example, the Lyman-break or H $\alpha$ , the abundance of observable Ly $\alpha$ -selected galaxies will be modulated by the presence of neutral hydrogen, impacting their clustering (Furlanetto et al. 2006c; McQuinn et al. 2007a, 2007b; Mesinger & Furlanetto 2008a, 2008b) and the 21 cm galaxy cross-power spectrum. This modulation occurs because damping wing absorption extinguishes the Ly $\alpha$  line for sources sufficiently close to the edge of an H II region (Miralda-Escudé 1998), where there is an adjacent column of neutral hydrogen. This means Ly $\alpha$ -selected galaxies will lie toward the center of rather large,  $R \gtrsim 1$  proper Mpc, H II regions (Furlanetto et al. 2006c; McQuinn et al. 2007a, 2007b; Mesinger & Furlanetto 2008a, 2008b). Owing to this, and because observable galaxies will have larger masses after Ly $\alpha$  selection, the clustering of Ly $\alpha$ -selected galaxies should increase as such galaxies are detected at earlier and earlier stages of reionization. Thus far our mock galaxies have been uniformly selected—i.e., not modulated by the presence of neutral hydrogen. While this is appropriate for Lyman-break-selected galaxies, it is incorrect for Ly $\alpha$ -selected galaxies before reionization completes.

In order to examine the impact of Ly $\alpha$  selection on the 21 cm galaxy cross-spectrum, we compute the damping wing optical



**Figure 10.** 21 cm galaxy cross-power spectrum for Ly $\alpha$ -selected galaxies. Top panel: the 21 cm galaxy cross-power spectrum for each of Ly $\alpha$ -selected galaxies (dashed lines) and “all galaxies” (solid lines). We show results at  $(\langle x_i \rangle, z) = (0.21, 8.34), (0.54, 7.32), (0.82, 6.90)$ , and  $M_{g,\min} = 10^{10} M_\odot$ . Bottom panel: the cross-correlation coefficient between the 21 cm and galaxy fields for the model curves in the top panel.

(A color version of this figure is available in the online journal.)

depth,  $\tau_D$ , toward each of our target halos. For simplicity, we calculate only the damping wing optical depth at line center (see e.g., Equations (1) and (2) of Mesinger & Furlanetto 2008b) and do not model resonant absorption (see McQuinn et al. 2007a; Mesinger & Furlanetto 2008a; Dijkstra et al. 2007 for discussion). Assuming that each source’s luminosity is proportional to its host-halo mass, and adopting our fiducial choice of  $M_{g,\min} = 10^{10} M_\odot$ , our Ly $\alpha$  survey detects sources with  $M \exp[-\tau_D] \geq M_{g,\min}$ .

With mock Ly $\alpha$ -selected galaxy catalogs in hand, we calculate the cross-power spectrum of these galaxies with the 21 cm field for a few outputs of differing ionization fractions and redshifts. The results of this calculation are shown in Figure 10. Comparing the cross spectra of the Ly $\alpha$ -selected galaxies (top panel, dashed lines) and uniformly selected galaxies (solid lines), we see that the large-scale amplitude of the cross spectra is higher, and that the signal turns over on larger scales, for the Ly $\alpha$ -selected galaxy samples. It is easy to understand these trends qualitatively. For a galaxy to be visible in Ly $\alpha$  it must reside in a sufficiently large H II region—or more accurately, it needs to reside along a sufficiently long ionized *skewer*—to avoid complete attenuation owing to damping wing absorption. The largest H II regions form around the most clustered sources, and so the galaxies detectable in Ly $\alpha$  are more clustered than uniformly selected galaxies of the same host-halo mass (Furlanetto et al. 2006c, McQuinn et al. 2007a, 2007b, Mesinger & Furlanetto 2008a, 2008b). This enhanced clustering is reflected in the boosted large-scale 21 cm galaxy cross-power spectrum. Likewise, the turnover on small scales is set by the characteristic H II region size around *detectable* galaxies, which increases for the Ly $\alpha$ -selected galaxies: Ly $\alpha$  galaxies residing in small bubbles are attenuated out of the sample by damping wing absorption.

This is visualized most clearly in the cross-correlation coefficient between the 21 cm and galaxy fields (Figure 10, bottom panel). In the uniformly selected galaxy sample, the correlation coefficient turns over on progressively larger scales as reionization proceeds. By contrast, the small-scale turnover in the Ly $\alpha$ -selected sample is relatively fixed. At early times when the bubbles are small, the turnover in the cross-spectrum with the Ly $\alpha$ -selected sample largely reflects the damping wing scale. In order to best characterize bubble growth in the early and middle stages of reionization, one requires a uniformly selected galaxy sample, rather than a Ly $\alpha$ -selected sample. Finally, note that the cross-correlation coefficient in the Ly $\alpha$ -selected samples does not quite reach  $r = -1$  on large scales. This behavior is enhanced early in reionization, and presumably results because small bubbles are missed by Ly $\alpha$  selection, which contribute most significantly at low-ionized fractions. In summary, while Ly $\alpha$ -selected samples will be interesting for initial cross-spectrum detections, uniformly selected samples will be required to best constrain bubble growth during reionization.

## 7. DETECTABILITY

In this section, we calculate the statistical significance at which future surveys can detect the 21 cm galaxy cross-spectrum and briefly consider the resulting insights into reionization. Here we closely follow the calculations in Furlanetto & Lidz (2007), by simply extending them to incorporate our simulated cross-spectrum signal. In our calculations we consider a 21 cm survey with the specifications planned for each of the MWA (Bowman et al. 2006; McQuinn et al. 2006; Mao et al. 2008) and LOFAR (Jelic et al. 2008), which we review below (Sections 7.2 and 7.3). We consider two basic types of galaxy surveys. First, we consider a survey similar to the Subaru deep-field survey for LAEs (Kashikawa et al. 2006). Since the Subaru survey is ongoing, this calculation should illustrate what is achievable in the near future as the MWA and LOFAR come online. Note that the present Subaru deep field does not overlap with the planned MWA target fields (M. Morales 2008, private communication), but our calculations still serve to illustrate what is possible in the near future. Next, we consider a more futuristic galaxy survey. Coupling our futuristic galaxy survey with the MWA or LOFAR, one can potentially measure the cross-power spectrum at several redshifts, probing the evolution in the cross-spectrum signal, and tracing the growth of H II regions during reionization as in Figure 4.

### 7.1. Statistical Error Estimates

To begin with, we describe our statistical error estimates, reviewing the formulae for cross-spectrum error bars for a survey of given specifications, incorporating sample variance, thermal noise in the 21 cm radio telescope, and shot-noise and redshift errors in the galaxy distribution. Here we restrict ourselves to the spherically averaged cross-spectrum, since the MWA and LOFAR have limited transverse sensitivity and since very precise galaxy redshifts will be required to measure the angular dependence of the cross-spectrum (Furlanetto & Lidz 2007).

We generally find it convenient to estimate error bars on the cross-correlation coefficient,  $r(k)$ , rather than on the cross-spectrum itself. We desire an estimate of the error bar on  $r(k)$  calculated from spherically averaged auto and cross spectra in a bin of logarithmic width  $\epsilon = d \ln k$ . For notational convenience let us denote the cross-correlation coefficient by



$r(k) = P_{21,\text{gal}}(k)/[P_{21}(k)P_{\text{gal}}(k)]^{1/2} = A(k)/[B(k)C(k)]^{1/2}$ . Propagating errors, the fractional error on the cross-correlation coefficient is

$$\frac{\sigma_r^2}{r^2}(k) = \frac{\sigma_A^2}{A^2}(k) + \frac{\sigma_B^2}{4B^2}(k) + \frac{\sigma_C^2}{4C^2}(k) - \frac{\sigma_{AB}^2}{AB}(k) - \frac{\sigma_{AC}^2}{AC}(k) + \frac{\sigma_{BC}^2}{2BC}(k). \quad (7)$$

This expression involves the cross-spectrum variance, the 21 cm and galaxy power spectrum variances, and the co-variance between the various power spectra, each calculated for spherically averaged power spectra in shells of logarithmic width  $\epsilon$ . The disadvantage of considering the cross-correlation coefficient is that the cross-spectrum can, under appropriate circumstances, be detected at higher sensitivity than the 21 cm autospectrum (Furlanetto & Lidz 2007). In this case, the error bar on the cross-correlation coefficient, which includes an error term from the autospectrum, will be larger than that for the cross-spectrum alone. Furthermore, estimating the cross-correlation coefficient requires an autospectrum estimate and is hence more susceptible to residual foreground contamination than the cross-spectrum alone.

Consider first the power spectrum variance terms for a single  $k$ -mode, with line of sight component  $k_{\parallel} = \mu k$ , restricting ourselves to modes in the upper-half plane. The power spectrum variance expressions are as follows (Furlanetto & Lidz 2007):

$$\sigma_A^2(k, \mu) = \text{var}[P_{21,\text{gal}}(k, \mu)] = \frac{1}{2} [P_{21,\text{gal}}^2(k, \mu) + \sigma_B(k, \mu)\sigma_C(k, \mu)], \quad (8)$$

$$\sigma_B^2(k, \mu) = \text{var}[P_{21}(k, \mu)] = \left[ P_{21}(k, \mu) + \frac{T_{\text{sys}}^2}{T_0^2} \frac{1}{B t_{\text{int}}} \frac{D^2 \Delta D}{n(k_{\perp})} \left( \frac{\lambda^2}{A_e} \right)^2 \right]^2, \quad (9)$$

$$\sigma_C^2(k, \mu) = \text{var}[P_{\text{gal}}(k, \mu)] = [P_{\text{gal}}(k, \mu) + n_{\text{gal}}^{-1} e^{k_{\parallel}^2 \sigma_z^2}]^2. \quad (10)$$

The second term in Equation (9) comes from thermal noise in the radio telescope, the second term in Equation (10) expresses the shot-noise error, while the other terms in the above equations are sample variance contributions. The thermal noise term depends on the system temperature,  $T_{\text{sys}}$ ; the comoving distance to the center of the survey at redshift  $z$ ,  $D(z)$ ; the survey depth,  $\Delta D$ ; the observed wavelength,  $\lambda$ ; the effective area of each antenna tile,  $A_e$ ; the survey bandwidth,  $B$ ; the total observing time,  $t_{\text{int}}$ ; and the distribution of antennas. The factor  $T_0$  in the denominator of the detector noise term arises because we normalize the 21 cm field by  $T_0$  so that it is dimensionless—i.e., we work with the field  $\delta_T/T_0$ . The dependence on antenna configuration is encoded in  $n(k_{\perp})$  which denotes the number density of baselines observing a mode with transverse wavenumber  $k_{\perp}$  (McQuinn et al. 2006; Bowman et al. 2006; Lidz et al. 2007b). The galaxy shot-noise term depends on  $n_{\text{gal}}$  which is the abundance of galaxies observable in our hypothetical survey, and on the accuracy of the galaxy redshifts obtained by the survey. The galaxy redshift error is given in comoving units by  $\sigma_x = c\sigma_z/H(z)$ .

We also require expressions for the covariance between the different power spectra. These can be computed straightforwardly as

$$\sigma_{AB}^2(k, \mu) = \text{cov}[P_{21,\text{gal}}(k, \mu), P_{21}(k, \mu)] = P_{21,\text{gal}}(k, \mu)P_{21}(k, \mu), \quad (11)$$

$$\sigma_{AC}^2(k, \mu) = \text{cov}[P_{21,\text{gal}}(k, \mu), P_{\text{gal}}(k, \mu)] = P_{21,\text{gal}}(k, \mu)P_{\text{gal}}(k, \mu), \quad (12)$$

$$\sigma_{BC}^2(k, \mu) = \text{cov}[P_{21}(k, \mu), P_{\text{gal}}(k, \mu)] = P_{21,\text{gal}}^2(k, \mu). \quad (13)$$

Finally, we can estimate the error bar on the cross-correlation coefficient formed from our spherically averaged power spectra. We do this by adding the power-spectrum error bars for individual  $k$ -modes from Equations (8)–(13) in inverse quadrature, performing a similar calculation for each individual term in Equation (7). For example, the variance of the cross-spectrum averaged over a spherical shell of the logarithmic width  $\epsilon = d \ln k$  is

$$\frac{1}{\sigma_A^2(k)} = \sum_{\mu} \frac{\epsilon k^3 V_{\text{survey}}}{4\pi^2} \frac{\Delta\mu}{\sigma_A^2(k, \mu)}. \quad (14)$$

The effective survey volume for our radio telescope is  $V_{\text{survey}} = D^2 \Delta D (\lambda^2/A_e)$ . If the galaxy survey has a lesser volume,  $V_{\text{gal}}$ , then the variance of the binned power spectrum estimated from this lesser volume (for a mode contained within the lesser survey volume) is larger by a factor of  $\sim V_{\text{gal}}/V_{\text{survey}}$ .

## 7.2. The MWA

With these expressions in hand let us briefly describe the specifications we assume for our 21 cm and galaxy surveys. The MWA will have a large field of view, spanning  $\sim 800 \text{ deg}^2$  on the sky, and consisting of 500 antenna tiles each with an effective area of  $A_e = 14 \text{ m}^2$  at  $z = 8$  (Bowman et al. 2006). Each antenna tile is 4m wide, and we follow Bowman et al. (2006) and McQuinn et al. (2006) in assuming that the antennas are packed as closely as possible within a compact core, with the distribution subsequently falling off as  $r^{-2}$  in order to capture large baselines, out to a maximum baseline of 1.5 km. Lidz et al. (2007b) argued that a compact antenna configuration, with all of the MWA's antennas packed as close as possible, is a superior configuration for 21 cm autospectrum measurements. This configuration is less good for the cross-spectrum: given a galaxy survey with photometric redshifts, one needs to balance the MWA's high line-of-sight sensitivity, yet poor transverse sensitivity, with the galaxy survey's high transverse sensitivity, yet poor line-of-sight sensitivity owing to redshift uncertainties.

We assume that the system temperature is set by the sky temperature, which we take to be  $T_{\text{sys}} = 280(1 + z/7.5)^{2.3} \text{ K}$ , following Wyithe & Morales (2007). We consider a bandwidth of  $B = 6 \text{ MHz}$  observing for a total time of  $t_{\text{int}} = 1000 \text{ h}$ . The bandwidth is chosen to be small enough to ensure that the signal evolves minimally over the corresponding redshift interval (McQuinn et al. 2006).

## 7.3. LOFAR

Initial estimates showed that LOFAR and the MWA should have comparable sensitivity for detecting the 21 cm autospectrum (McQuinn et al. 2006; Mao et al. 2008). In these initial estimates, LOFAR had a smaller field of view than the MWA (by a factor of  $\gtrsim 10$ ), but it had a larger collecting area, which compensated for the reduced sky coverage. Since the larger field of view of the MWA is wasted when cross correlating with a galaxy survey that covers a much smaller patch

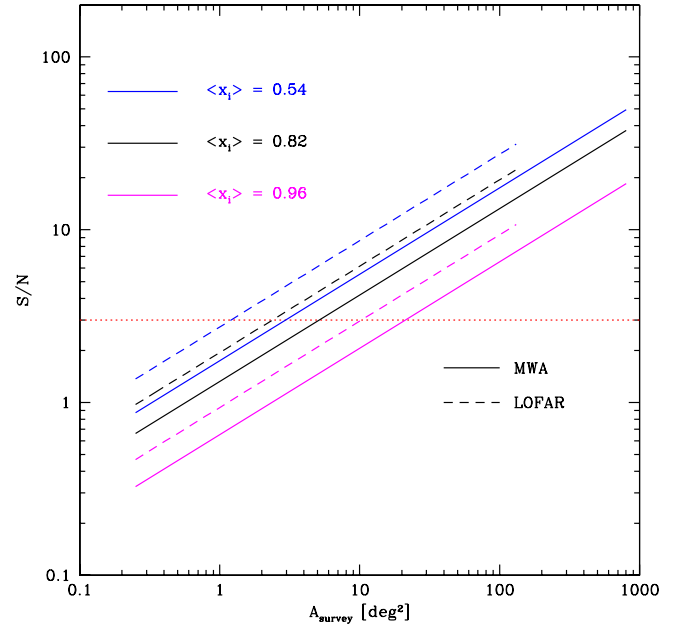
on the sky, one expected LOFAR to be more sensitive than the MWA for detecting the 21 cm galaxy cross-spectrum (Furlanetto & Lidz 2007). Unfortunately, budget cuts have modified the LOFAR specifications and its resulting sensitivity for detecting the 21 cm auto and cross spectra, and so we need to double check this expectation. Specifically, the new LOFAR EoR core has reduced collecting area, but covers a larger field of view on the sky, has more baselines, and a smaller minimum baseline.

Here we consider the sensitivity of the new LOFAR EoR core. The precise LOFAR specifications are still evolving, but we follow Jelic et al. (2008, their Section 6) and the simple model of McQuinn et al. (2006) as a plausible estimate. According to the new specifications, LOFAR will consist of 48 large antenna substations distributed within a 1 km radius, with a minimum baseline of 40 m. Each substation consists of  $N_{\text{dipole}} = 24 \times 16$  dual-polarization dipole antennas. In order to calculate the effective area of each substation, we follow the estimates of Bowman et al. (2006) for the effective area of the MWA tiles, and assume that each LOFAR substation has an effective area of  $A_e \sim N_{\text{dipole}} \lambda^2 / 4$  for wavelengths below  $\lambda \sim 2$  m. This estimate neglects shadowing effects and assumes independent, perfect dipoles above an infinite, conducting ground screen.<sup>7</sup> Each LOFAR station simultaneously observes four separate regions on the sky. For the simple estimates here, we assume that LOFAR's antenna stations are closely packed in a compact core, before tapering off in an  $r^{-2}$  configuration out to a maximum radius of 1 km. As for the MWA, we consider 1000 h. of LOFAR observations over a bandwidth of  $B = 6$  MHz.

#### 7.4. Subaru-like Survey

We first consider the detectability of the 21 cm galaxy cross-spectrum obtainable by combining the MWA and LOFAR with the Subaru deep-field survey, and plausible extensions. The existing Subaru deep-field survey has a  $0.25 \text{ deg}^2$  field of view and locates LAEs near  $z = 6.6$  to a depth of  $130 \text{ \AA}$ . The existing spectroscopically confirmed Subaru deep-field sample at redshift  $z = 6.6$  consists of 36 emitters (Kashikawa et al. 2006). The number density of spectroscopically confirmed emitters corresponds to  $n_{\text{gal}} = 1.6 \times 10^{-4} \text{ Mpc}^{-3}$ . An extension to the Subaru deep field, the Subaru/XMM-Newton Deep Survey is already underway, and promises to increase the observed  $z = 6.6$  field of view by a factor of  $\sim 4$ , reaching a survey area of  $A_{\text{survey}} \sim 1 \text{ deg}^2$  by the end of the year (Ouchi et al. 2005).

Given the rapid progress in the area surveyed, we examine how the detectability of the cross-spectrum scales with increasing field-of-view, at a fixed depth and a galaxy number density. In practice, we calculate the cross-spectrum signal-to-noise ratio (S/N) for a galaxy survey that covers the full field-of-view of the MWA ( $\sim 800 \text{ deg}^2$  at this redshift), and scale the S/N (squared) in each  $k$ -bin downward by the ratio of the galaxy survey volume to that of the MWA (see Equation (14)). We perform a similar calculation for LOFAR. For each model cross-spectrum, our S/N estimates assume a galaxy number density of  $n_{\text{gal}} = 1.6 \times 10^{-4} \text{ Mpc}^{-3}$ , and redshift errors for the spectroscopically confirmed galaxies of  $\sigma_z = 0.01$ . The assumed redshift error corresponds to a velocity of several hundred  $\text{km s}^{-1}$ , motivated by the typical velocity offsets for Ly $\alpha$  lines observed by Shapley et al. (2003) in Lyman-break galaxies at  $z \sim 3$ . The total S/N (squared) is determined by summing the



**Figure 11.** The S/N for cross-spectrum detection. The S/N at which Subaru-like surveys, coupled with the MWA (solid lines) and LOFAR (dashed lines), can detect the 21 cm galaxy cross-spectrum as a function of survey area at  $z = 6.6$ . The different curves indicate different models for the ionization fraction. Each curve extends from the current Subaru area ( $\sim 0.25 \text{ deg}^2$ ) to the full field of view of MWA ( $\sim 820 \text{ deg}^2$ ) or LOFAR ( $\sim 140 \text{ deg}^2$  for the new specifications after observing four separate regions of the sky). The red dotted line indicates a  $3\text{-}\sigma$  detection of the cross-spectrum.

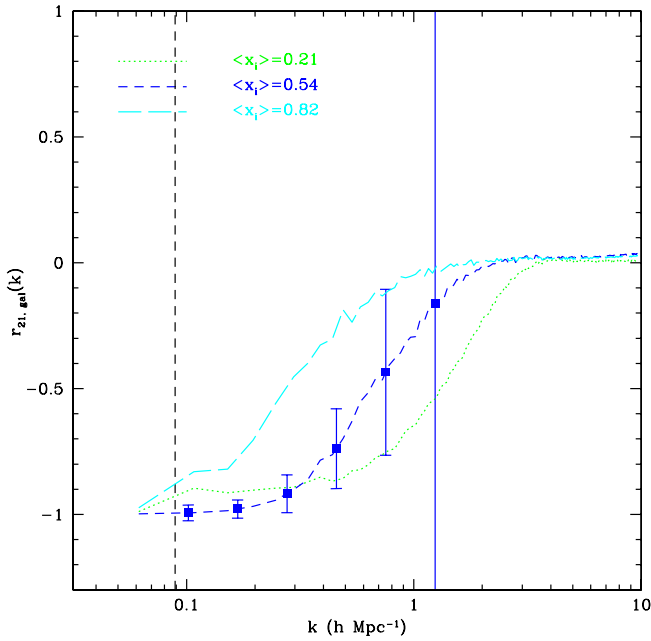
(A color version of this figure is available in the online journal.)

signal squared divided by our variance estimate (Equations 8 and 14) over all detected  $k$ -bins.

In Figure 11 we show the detectability of the cross-spectrum for a few different models over a range of survey areas. In each model we adopt a plausible minimum detectable galaxy mass of  $M_{\text{g,min}} = 10^{10} M_{\odot}$ , fixing the galactic duty cycle to match the observed Subaru deep-field abundance of  $n_{\text{gal}} = 1.6 \times 10^{-4} \text{ Mpc}^{-3}$ . The corresponding duty cycle in our models is around  $\sim 1\%$ . Given that we currently have few direct observational constraints on the filling factor of H II regions near  $z = 6.6$ , we consider models in which the ionized fraction is  $\langle x_i \rangle = 0.54, 0.82$ , and  $0.96$  at this redshift. Strictly speaking, we should use our Ly $\alpha$ -selected cross-spectrum models here, but at these ionization fractions we expect this to boost our S/N only slightly (Figure 10). For simplicity, we conservatively ignore the clustering boost from Ly $\alpha$  selection here. The amplitude of the cross-spectrum is largest amongst these models at  $\langle x_i \rangle = 0.54$ , and is substantially smaller by  $\langle x_i \rangle = 0.96$  (see Figure 4), and so the more neutral models will be easier to detect.

The results shown in Figure 11 illustrate that cross-correlating the MWA with a galaxy survey of a size comparable to the present Subaru deep-field survey will not allow a significant cross-spectrum detection ( $S/N \lesssim 1\sigma$ ), even if the IGM is significantly neutral at  $z = 6.6$ . However, extensions to the Subaru deep field that cover a larger area on the sky should yield significant cross-spectrum detections. For example, extending the present sky coverage by a factor of  $\sim 10\text{--}15$  to  $3 \text{ deg}^2$  should provide a  $\gtrsim 2\text{--}3\sigma$  cross-spectrum detection in our  $\langle x_i \rangle = 0.54$  and  $\langle x_i \rangle = 0.82$  models, but only a  $\sim 1\sigma$  detection in our  $\langle x_i \rangle = 0.96$  model. The S/N for cross-correlating with the new version of LOFAR is still a bit better than that for cross correlating with the MWA. The fact that LOFAR has a

<sup>7</sup> The effective area per LOFAR tile in McQuinn et al. (2006) and Mao et al. (2008) is a factor of 2 more pessimistic than the area from this formula.



**Figure 12.** The error estimate for the 21 cm galaxy cross-correlation coefficient. Here we consider a futuristic galaxy survey covering the entire MWA field of view, cross-correlated with the MWA. The blue points show the mean signal and error estimates for our hypothetical 21 cm/galaxy survey when  $\langle x_i \rangle = 0.54$ . The other curves show the cross-correlation coefficient when  $\langle x_i \rangle = 0.21$  and  $0.82$  respectively. Our hypothetical survey should help constrain the volume-weighted ionization fraction. The vertical black dashed line shows the wave number corresponding to the survey depth, below which foreground contamination will prohibit extracting the signal.

(A color version of this figure is available in the online journal.)

smaller field of view than the MWA ( $\sim 140 \text{ deg}^2$  after observing four separate regions on the sky, compared to  $\sim 800 \text{ deg}^2$ ), is unimportant when cross correlating with a galaxy survey that spans a smaller field of view. The revised LOFAR still has  $\sim 2\text{--}2.5$  times more collecting area than the MWA by our estimates, which gives it a superior cross-spectrum sensitivity. (Note that we assume a factor of  $\sim 2$  larger collecting area per LOFAR dipole than in McQuinn et al. 2006 and Mao et al. 2008, with our current numbers being motivated by the  $A_e = N_{\text{dipole}} \lambda^2 / 4$  formula from Bowman et al. 2006.) Cross correlating LOFAR with a galaxy survey of  $2\text{--}3 \text{ deg}^2$  should allow a  $\gtrsim 3\text{-}\sigma$  cross-spectrum detection in our  $\langle x_i \rangle = 0.54$  and  $\langle x_i \rangle = 0.82$  models. As mentioned earlier, the Subaru survey may reach this sky coverage relatively soon, making a cross-spectrum detection feasible in the next few years if the IGM is partly neutral around  $z \sim 6.6$ . More ambitious surveys covering the entire MWA sky  $\sim 800 \text{ deg}^2$  would clearly move beyond mere detections—the detection S/N for such surveys is at the tens of sigma level (see Figure 11)—and provide valuable constraints on reionization models.

### 7.5. Futuristic Survey

Since more futuristic surveys will go beyond mere detections, we proceed to consider the constraining power of a large field-of-view galaxy survey—cross correlated with the MWA—in more detail. Futuristic surveys will allow one to probe small scales and capture the turnover in the cross-correlation coefficient. We calculate the expected error bar on the cross-correlation coefficient as a function of wave number for a galaxy survey spanning the full MWA field of view, and consider the ability of this survey to constrain reionization models. Here we assume

that the galaxy survey can detect fainter galaxies, reaching a galactic abundance 100 times larger than in the previous section, with the same redshift accuracy of  $\sigma_z = 0.01$ . We consider a redshift of  $z = 7.3$ .

Using again the models of Section 3 as input, we estimate the statistical sensitivity of our futuristic galaxy survey. The results of our sensitivity calculation are shown in Figure 12, for spherical bins of logarithmic width  $\epsilon = 0.5$ . Here we plot the simulated signal when the IGM is  $\sim 50\%$  ionized along with a statistical error estimate for our hypothetical survey. For contrast, we additionally show theoretical model curves when the IGM is each of  $\sim 20\%$  and  $\sim 80\%$  ionized.

The curves and error bars in Figure 12 show that the statistical precision of our hypothetical survey is high enough to distinguish between the different stages of reionization shown in the figure over about a decade in scale. On large scales, the measurement is limited by foreground removal while on small scales 21 cm detector noise and galaxy redshift errors limit the statistical precision of the measurements (Furlanetto & Lidz 2007). Although the 21 cm galaxy cross-power spectrum signal is much less susceptible than the 21 cm auto-power spectrum to foreground contamination, free-free and synchrotron emission from the high-redshift galaxies in our survey still contaminate the 21 cm galaxy cross-power spectrum somewhat (Furlanetto & Lidz 2007). This prohibits measuring modes with lines of sight wave numbers  $k_{\parallel} < 2\pi/\Delta D$ , where  $\Delta D$  is the depth of the survey. The discreteness of the survey means that the only modes in our survey that satisfy this requirement have  $k_{\parallel} = 0$ ; hence all modes with  $k \leq k_{\parallel}$  will be removed in the foreground cleaning process. The black dashed line indicates the wavenumber corresponding to the survey depth in our hypothetical survey. We should remind the reader here of one trade-off involved with considering the cross-correlation coefficient rather than the cross-spectrum alone. The cross-correlation coefficient is a more convenient quantity than the cross-spectrum for visualizing the small-scale turnover (see Figure 12), but it is less desirable in that it includes the autospectrum, which is more susceptible to foreground contamination.

The sensitivity estimates shown in Figure 12 are encouraging, and suggest that future 21 cm galaxy surveys may help constrain the filling factor and size distribution of H II regions during reionization. Comparing our error estimates with the results of Figure 9 suggests that futuristic surveys might also—by measuring the cross-spectrum in different galaxy luminosity bins—weakly constrain the dependence of bubble size on host-halo mass. Note also that the thermal noise term in the 21 cm variance (see Equation (9)) still contributes significantly for most  $k$ -bins shown here, and so futuristic 21 cm surveys with more antennas and larger collecting areas than the MWA can further improve cross-spectrum sensitivity. In particular, a future FFT telescope (Tegmark & Zaldarriaga 2008) should boost the sensitivity compared to our estimates here (see Mao et al. 2008 for estimates of the autospectrum sensitivity with an FFT telescope).

## 8. CONCLUSIONS

In this paper, we considered the scientific return of future 21 cm galaxy cross-power spectrum measurements. A strong cross-spectrum measurement ultimately requires detecting a sizable number of high-redshift galaxies over a large field of view, which presents a significant observational challenge. Nonetheless, we showed that a detection of the cross-spectrum may be achieved in the near future by combining LOFAR or the



MWA and the Subaru survey for LAEs at  $z \sim 6.6$ . We estimate that a  $\sim 3\sigma$  detection is feasible, provided the IGM is  $\gtrsim 20\%$  neutral at this redshift, and that the Subaru survey's sky coverage is extended from  $0.25 \text{ deg}^2$  to  $\sim 2\text{--}3 \text{ deg}^2$ . This detection would already be quite valuable, as it would help confirm that the detected 21 cm signal comes from the high-redshift IGM, and not from foreground contamination, which should mostly be uncorrelated with high-redshift galaxies (Furlanetto & Lidz 2007). According to our estimates, LOFAR is slightly more sensitive than the MWA for detecting the 21 cm galaxy cross-spectrum in conjunction with mild extensions to the Subaru survey.

Futuristic galaxy surveys covering 100 s of square degrees on the sky, can be combined with the MWA, LOFAR, and other 21 cm surveys, to move beyond a mere detection of the cross-spectrum signal and map out its detailed scale dependence. The galaxy surveys required for these measurements are clearly very challenging, but rapid progress is being made in this direction as deep, wide field surveys are being designed to study baryonic acoustic oscillations and/or weak lensing at high redshift (e.g., ADEPT, HETDEX,<sup>8</sup> CIP, and others). Another option is to sparsely sample the MWA or LOFAR fields, in order to capture the large-scale modes (Furlanetto & Lidz 2007). We have shown that the 21 cm galaxy cross-spectrum is a relatively direct tracer of bubble growth during reionization. Measuring the turnover scale as a function of galaxy luminosity constrains the luminosity dependence of the characteristic bubble size. This information is difficult, or impossible, to obtain with the 21 cm autospectrum alone. In order to extract the most information out of the cross-spectrum, it should be combined with measurements of the galaxy autospectrum and luminosity function, which will help to constrain the galaxy-luminosity-halo-mass correlation.

A further interesting feature of the simulated signal is that the cross-correlation changes sign on large scales near the beginning of reionization (Figure 4). At this early phase of reionization, our results may, however, be modified by spin temperature fluctuations, which we presently neglect. Future work should incorporate these fluctuations. If our signature holds up, the change in sign of the cross-correlation would provide a very interesting observational indicator of the earliest phases of reionization. Finally, we found that the 21 cm galaxy cross-power spectrum might provide an interesting observational signature of scenarios where ionizing photons fail to escape from low-mass halos. Provided galaxies in these low-mass halos are detectable longward of the ionization edge, we expect the cross-spectrum to change sign and turn positive on small scales. Generally speaking, the 21 cm galaxy cross-spectrum is a more direct tracer of the impact of galaxies on the surrounding IGM than the 21 cm autospectrum. As such, it can potentially provide a wealth of information about the EoR and early structure formation.

We thank Mark Dijkstra and Miguel Morales for helpful discussions. We thank Suvendra Dutta for useful conversations and for his collaboration in related work. We are grateful to Saleem Zaroubi and the anonymous referee for pointing us to the new LOFAR specifications in Jelic et al. (2008), and to Saleem Zaroubi for helpful conversations regarding LOFAR. Support was provided, in part, by the David and Lucile Packard Foundation, the Alfred P. Sloan Foundation, and grants

AST-0506556 and NNG05GJ40G. O.Z. acknowledges additional support by a Berkeley Center for Cosmological Physics (BCCP) Fellowship.

## REFERENCES

- Abel, T., & Wandelt, B. D. 2002, *MNRAS*, **330**, 53  
 Barkana, R., & Loeb, A. 2001, *Phys. Rep.*, **349**, 125  
 Barkana, R., & Loeb, A. 2002, *ApJ*, **578**, 1  
 Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2008, *ApJ*, **686**, 230  
 Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, *ApJ*, **638**, 20  
 Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, *Phys. Rev. Lett.*, **100**, 091303  
 Ciardi, B., & Madau, P. 2003, *ApJ*, **596**, 1  
 Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, *MNRAS*, **373**, 369, 05  
 Dijkstra, M., Haiman, Z., Rees, M. J., & Weinberg, D. H. 2004, *ApJ*, **601**, 666  
 Dijkstra, M., Lidz, A., & Wyithe, J. S. B. 2007, *MNRAS*, **377**, 1175  
 Furlanetto, S. R., & Lidz, A. 2007, *ApJ*, **660**, 1030  
 Furlanetto, S. R., McQuinn, M., & Hernquist, L. 2006a, *MNRAS*, **365**, 115  
 Furlanetto, S. R., & Oh, S. P. 2005, *MNRAS*, **363**, 1031  
 Furlanetto, S. R., Oh, S. P., & Briggs, F. 2006b, *Phys. Rep.*, **433**, 181  
 Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *ApJ*, **613**, 1  
 Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2006c, *MNRAS*, **365**, 1012  
 Gnedin, N. Y., Kravtsov, A. V., & Chen, H.-W. 2008, *ApJ*, **672**, 765  
 Haiman, Z., Abel, T., & Madau, P. 2001, *ApJ*, **551**, 599  
 Jelic, V., et al. 2008, *MNRAS*, **389**, 1319  
 Kashikawa, N., et al. 2006, *ApJ*, **648**, 7  
 Komatsu, E., et al. 2008, *ApJS*, submitted, arXiv:0803.0547  
 Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., Dutta, S., & Hernquist, L. 2007a, *ApJ*, **659**, 865  
 Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2007b, *ApJ*, **680**, 962  
 Madau, P., Meiksin, A., & Rees, M. J. 1997, *ApJ*, **475**, 429  
 Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, *Phys. Rev. D*, **78**, 023529  
 McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007a, *MNRAS*, **381**, 75  
 McQuinn, M., Lidz, A., Zahn, O., Dutta, S., Hernquist, L., & Zaldarriaga, M. 2007b, *MNRAS*, **377**, 1043  
 McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, *ApJ*, **653**, 815  
 Mesinger, A., & Furlanetto, S. R. 2007, *ApJ*, **669**, 663  
 Mesinger, A., & Furlanetto, S. R. 2008a, *MNRAS*, **386**, 1990  
 Mesinger, A., & Furlanetto, S. R. 2008b, *MNRAS*, **385**, 1348  
 Miralda Escudé, J. 1998, *ApJ*, **501**, 15  
 Miralda Escudé, J., Haehnelt, M., & Rees, M. J. 2000, *ApJ*, **530**, 1  
 Navarro, J. F., & Steinmetz, M. 1997, *ApJ*, **478**, 13  
 Oh, S. P., & Haiman, Z. 2003, *MNRAS*, **346**, 456  
 Ouchi, M., et al. 2005, *ApJ*, **620**, L1  
 Partridge, R. B., & Peebles, P. J. E. 1967, *ApJ*, **147**, 868  
 Pritchard, J. R., & Furlanetto, S. R. 2007, *MNRAS*, **376**, 1680  
 Pritchard, J. R., & Loeb, A. 2008, *Phys. Rev. D*, submitted, arXiv:0802.2102  
 Rhoads, J. E., et al. 2004, *ApJ*, **611**, 59  
 Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, *ApJ*, **546**, 20  
 Scott, D., & Rees, M. J. 1990, *MNRAS*, **247**, 510  
 Shapiro, P. R., Iliev, I. T., & Raga, A. C. 2004, *MNRAS*, **348**, 753  
 Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, **588**, 65  
 Sokasian, A., Abel, T., & Hernquist, L. 2001, *New Astron.*, **6**, 359  
 Sokasian, A., Abel, T., Hernquist, L., & Springel, V. 2003, *MNRAS*, **344**, 607  
 Springel, V. 2005, *MNRAS*, **364**, 1105  
 Springel, V., & Hernquist, L. 2003, *MNRAS*, **339**, 312  
 Spergel, D. N., et al. 2007, *ApJS*, **170**, 377  
 Stark, D. P., Ellis, R. S., Richard, J., Kneib, J.-P., Smith, G. P., & Santos, M. R. 2007, *ApJ*, **663**, 10  
 Tegmark, M., & Zaldarriaga, M. 2008, *Phys. Rev. D*, submitted, arXiv:0805.4414  
 Thoul, A. A., & Weinberg, D. H. 1996, *ApJ*, **465**, 608  
 Wyithe, J. S. B., & Loeb, A. 2007, *MNRAS*, **375**, 1034  
 Wyithe, J. S. B., & Loeb, A. 2008, *MNRAS*, submitted, arXiv:0708.3392  
 Wyithe, J. S. B., & Morales, M. 2007, *MNRAS*, submitted, astro-ph/0703070  
 Zahn, O., Lidz, A., McQuinn, M., Dutta, S., Hernquist, L., Zaldarriaga, M., & Furlanetto, S. R. 2007, *ApJ*, **654**, 12  
 Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, *ApJ*, **608**, 622

<sup>8</sup> <http://www.as.utexas.edu/hetdex/>